

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
23 May 2002 (23.05.2002)

PCT

(10) International Publication Number  
**WO 02/40715 A2**

(51) International Patent Classification<sup>7</sup>: C12Q 1/68

(21) International Application Number: PCT/US01/27628

(22) International Filing Date:  
6 September 2001 (06.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/230,517	6 September 2000 (06.09.2000)	US
60/230,599	6 September 2000 (06.09.2000)	US
60/230,514	6 September 2000 (06.09.2000)	US
60/230,598	6 September 2000 (06.09.2000)	US
60/230,988	6 September 2000 (06.09.2000)	US
60/230,518	6 September 2000 (06.09.2000)	US
60/230,515	6 September 2000 (06.09.2000)	US
60/230,610	6 September 2000 (06.09.2000)	US
60/230,597	6 September 2000 (06.09.2000)	US
60/230,505	6 September 2000 (06.09.2000)	US
60/230,519	6 September 2000 (06.09.2000)	US
60/230,595	6 September 2000 (06.09.2000)	US
60/230,865	6 September 2000 (06.09.2000)	US
60/230,989	6 September 2000 (06.09.2000)	US
60/231,167	7 September 2000 (07.09.2000)	US
60/231,163	7 September 2000 (07.09.2000)	US
60/230,951	7 September 2000 (07.09.2000)	US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): JACKSON, Stuart [US/US]; 4913 Deep Creek Road, Fremont, CA 94555 (US). LINCOLN, Stephen, E. [US/US]; 725 Sapphire Street, Redwood City, CA 94061 (US). ALTUS, Christina, M. [US/US]; 625 Virginia Avenue, Campbell, CA 95008 (US). DUFOUR, Gerard, E. [US/US]; 5327 Greenridge Road, Castro Valley, CA 94552 (US). CHALUP, Michael, S. [US/US]; 183 Acalanes Drive, Apt. 6, Sunnyvale, CA 94086 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive, #17, Mountain View, CA 94040 (US). JONES, Anissa, Lee [US/US];

445 South 15th Street, San Jose, CA 95112 (US). YU, Jimmy, Y. [US/US]; 37330 Portico Terrace, Fremont, CA 94536 (US). WRIGHT, Rachel, J. [NZ/US]; 333 Anna Avenue, Mountain View, CA 94043 (US). GIETZEN, Darryl [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). LIU, Tommy, F. [US/US]; 201 Otilia Street, Daly City, CA 94014 (US). YAP, Pierre, E. [US/US]; 201 Happy Hollow Court, Lafayette, CA 94549 (US). DAHL, Christopher, R. [US/US]; 41277 Roberts Avenue, #6, Fremont, CA 94538 (US). MOMIYAMA, Monika, G. [US/US]; 25689 Deer Trail Place, Hayward, CA 94541 (US). BRADLEY, Diana, L. [US/US]; 3260 Crystal Heights Drive, Soquel, CA 95073 (US). ROHATGI, Sameer, D. [US/US]; 5 Rico Way #103, San Francisco, CA 94123 (US). HARRIS, Bernard [US/US]; 1014 Lupine Drive, Sunnyvale, CA 94086 (US). ROSEBERRY, Ann, M. [US/US]; 725 Sapphire Street, Redwood City, CA 94061 (US). GERSTIN, Edward, H., Jr. [US/US]; 1408 38th Avenue, San Francisco, CA 94122 (US). PERALTA, Careyna, H. [US/US]; 4585 Lakeshore Drive, Santa Clara, CA 95054 (US). DAVID, Marie, H. [US/US]; 131 Mirada Drive, Daly City, CA 94015 (US). PANZER, Scott, R. [US/US]; 965 East El Camino, #621, Sunnyvale, CA 94087 (US). FLORES, Vincent [US/US]; 35000 Begonia Street, Union City, CA 94587 (US). DAFFO, Abel [US/US]; 1750 Stokes St., #70, San Jose, CA 95126 (US). MARWAHA, Rakesh [CA/US]; 16272 Saratoga Street, #4, San Leandro, CA 94578 (US). CHEN, Alice, J. [US/US]; 4405 Norwalk Drive, #22, San Jose, CA 95129 (US). CHANG, Simon, C. [US/US]; 1901 Rock Street #103, Mountain View, CA 94043 (US). AU, Alan, P. [US/US]; 565 Ortega Avenue, #3, Mountain View, CA 94040 (US). INMAN, Rebekah, R. [US/US]; 210 Calderon Avenue, #23, Mountain View, CA 94041 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

[Continued on next page]

(54) Title: MOLECULES FOR DISEASE DETECTION AND TREATMENT

(57) Abstract: The present invention provides purified disease detection and treatment molecule polynucleotides (mddt). Also encompassed are the polypeptides (MDDT) encoded by mddt. The invention also provides for the use of mddt, or complements, oligonucleotides, or fragments thereof in diagnostic assays. The invention further provides for vectors and host cells containing mddt for the expression of MDDT. The invention additionally provides for the use of isolated and purified MDDT to induce antibodies and to screen libraries of compounds and the use of anti-MDDT antibodies in diagnostic assays. Also provided are microarrays containing mddt and methods of use.



WO 02/40715 A2

but many more may exist that are yet to be discovered.

DNA-based arrays can provide a simple way to explore the expression of a single polymorphic gene or a large number of genes. When the expression of a single gene is explored, DNA-based arrays are employed to detect the expression of specific gene variants. For example, a p53 tumor suppressor gene array is used to determine whether individuals are carrying mutations that predispose them to cancer. A cytochrome p450 gene array is useful to determine whether individuals have one of a number of specific mutations that could result in increased drug metabolism, drug resistance or drug toxicity.

DNA-based array technology is especially relevant for the rapid screening of expression of a large number of genes. There is a growing awareness that gene expression is affected in a global fashion. A genetic predisposition, disease or therapeutic treatment may affect, directly or indirectly, the expression of a large number of genes. In some cases the interactions may be expected, such as when the genes are part of the same signaling pathway. In other cases, such as when the genes participate in separate signaling pathways, the interactions may be totally unexpected. Therefore, DNA-based arrays can be used to investigate how genetic predisposition, disease, or therapeutic treatment affects the expression of a large number of genes.

The discovery of new molecules for disease detection and treatment satisfies a need in the art by providing new compositions which are useful in the diagnosis, study, prevention, and treatment of diseases associated with, as well as effects of exogenous compounds on, the expression of molecules for disease detection and treatment.

## SUMMARY OF THE INVENTION

The present invention relates to human disease detection and treatment molecule polynucleotides (mddt) as presented in the Sequence Listing. The mddt uniquely identify genes encoding structural, functional, and regulatory disease detection and treatment molecules.

The invention provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; c) a polynucleotide complementary to the polynucleotide of a); d) a polynucleotide complementary to the polynucleotide of b); and e) an RNA equivalent of a) through d). In one alternative, the polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252. In another alternative, the polynucleotide comprises at least 30 contiguous nucleotides of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; b) a polynucleotide comprising a naturally

occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; c) a polynucleotide complementary to the polynucleotide of a); d) a polynucleotide complementary to the polynucleotide of b); and e) an RNA equivalent of a) through d). In another alternative, the polynucleotide comprises at least 60 contiguous nucleotides of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; b) a polynucleotide comprising a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; c) a polynucleotide complementary to the polynucleotide of a); d) a polynucleotide complementary to the polynucleotide of b); and e) an RNA equivalent of a) through d). The invention further provides a composition for the detection of expression of disease detection and treatment molecule polynucleotides comprising at least one isolated polynucleotide comprising a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; c) a polynucleotide complementary to the polynucleotide of a); d) a polynucleotide complementary to the polynucleotide of b); and e) an RNA equivalent of a) through d); and a detectable label.

The invention also provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a polynucleotide sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence of a polynucleotide selected from the group consisting of SEQ ID NO:1-252; b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; c) a polynucleotide complementary to the polynucleotide of a); d) a polynucleotide complementary to the polynucleotide of b); and e) an RNA equivalent of a) through d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention also provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; c) a polynucleotide complementary to the polynucleotide of a); d) a polynucleotide complementary to the polynucleotide of b); and e) an RNA equivalent of a) through d). The method comprises a) hybridizing the sample with

a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide, and b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof. In one alternative, the invention provides a composition comprising a target polynucleotide of the method, wherein said probe comprises at least 30 contiguous nucleotides. In one alternative, the invention provides a composition comprising a target polynucleotide of the method, wherein said probe comprises at least 60 contiguous nucleotides.

The invention further provides a recombinant polynucleotide comprising a promoter sequence operably linked to an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; c) a polynucleotide complementary to the polynucleotide of a); d) a polynucleotide complementary to the polynucleotide of b); and e) an RNA equivalent of a) through d). In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a disease detection and treatment polypeptide, the method comprising a) culturing a cell under conditions suitable for expression of the disease detection and treatment polypeptide, wherein said cell is transformed with a recombinant polynucleotide, said recombinant polynucleotide comprising an isolated polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; iii) a polynucleotide complementary to the polynucleotide of i); iv) a polynucleotide complementary to the polynucleotide of ii); and v) an RNA equivalent of i) through iv), and b) recovering the disease detection and treatment polypeptide so expressed. The invention additionally provides a method wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:253-506.

The invention also provides an isolated disease detection and treatment polypeptide (MDDT) encoded by at least one polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252. The invention further provides a method of screening for a test compound that specifically binds to the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506. The method comprises a) combining the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506 with at least one

test compound under suitable conditions, and b) detecting binding of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506 to the test compound, thereby identifying a compound that specifically binds to the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506.

5           The invention further provides a microarray wherein at least one element of the microarray is an isolated polynucleotide comprising at least 30 contiguous nucleotides of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group  
10           consisting of SEQ ID NO:1-252; c) a polynucleotide complementary to the polynucleotide of a); d) a polynucleotide complementary to the polynucleotide of b); and e) an RNA equivalent of a) through d). The invention also provides a method for generating a transcript image of a sample which contains polynucleotides. The method comprises a) labeling the polynucleotides of the sample, b) contacting the elements of the microarray with the labeled polynucleotides of the sample under conditions suitable  
15           for the formation of a hybridization complex, and c) quantifying the expression of the polynucleotides in the sample.

          Additionally, the invention provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide  
20           sequence selected from the group consisting of SEQ ID NO:1-252; b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; c) a polynucleotide complementary to the polynucleotide of a); d) a polynucleotide complementary to the polynucleotide of b); and e) an RNA equivalent of a) through d). The method comprises a) exposing a sample comprising the target  
25           polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

          The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b)  
30           hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; iii) a polynucleotide  
35           complementary to the polynucleotide of i); iv) a polynucleotide complementary to the polynucleotide

of ii); and v) an RNA equivalent of i) through iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence  
5 selected from the group consisting of SEQ ID NO:1-252; ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252; iii) a polynucleotide complementary to the polynucleotide of i); iv) a polynucleotide complementary to the polynucleotide of ii); and v) an RNA equivalent of i) through iv), and alternatively, the target polynucleotide comprises a polynucleotide sequence of a  
10 fragment of a polynucleotide selected from the group consisting of i-v above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

15 The invention further provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group  
20 consisting of SEQ ID NO:253-506, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506. In one alternative, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:253-506.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from  
25 the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, and d) an immunogenic fragment of a  
30 polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506. In one alternative, the polynucleotide encodes a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:253-506. In another alternative, the polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-252.

35 Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence

selected from the group consisting of SEQ ID NO:253-506, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506.

The invention further provides a composition comprising a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid



sequence selected from the group consisting of SEQ ID NO:253-506, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional MDDT, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

#### DESCRIPTION OF THE TABLES

Table 1 shows the sequence identification numbers (SEQ ID NO:s) and template identification numbers (template IDs) corresponding to the polynucleotides of the present invention, along with the sequence identification numbers (SEQ ID NO:s) and open reading frame identification numbers (ORF IDs) corresponding to polypeptides encoded by the template ID.

Table 2 shows the sequence identification numbers (SEQ ID NO:s) and template identification numbers (template IDs) corresponding to the polynucleotides of the present invention, along with their GenBank hits (GI Numbers), probability scores, and functional annotations corresponding to the GenBank hits.

Table 3 shows the sequence identification numbers (SEQ ID NO:s) and template identification numbers (template IDs) corresponding to the polynucleotides of the present invention, along with polynucleotide segments of each template sequence as defined by the indicated "start" and "stop"

nucleotide positions. The reading frames of the polynucleotide segments and the Pfam hits, Pfam descriptions, and E-values corresponding to the polypeptide domains encoded by the polynucleotide segments are indicated.

5 Table 4 shows the sequence identification numbers (SEQ ID NO:s) and template identification numbers (template IDs) corresponding to the polynucleotides of the present invention, along with polynucleotide segments of each template sequence as defined by the indicated "start" and "stop" nucleotide positions. The reading frames of the polynucleotide segments are shown, and the polypeptides encoded by the polynucleotide segments constitute either signal peptide (SP) or transmembrane (TM) domains, as indicated. The membrane topology of the encoded polypeptide  
10 sequence is indicated, the N-terminus (N) listed as being oriented to either the cytosolic (N in) or non-cytosolic (N out) side of the cell membrane or organelle.

Table 5 shows the sequence identification numbers (SEQ ID NO:s) and template identification numbers (template IDs) corresponding to the polynucleotides of the present invention, along with component sequence identification numbers (component IDs) corresponding to each template. The  
15 component sequences, which were used to assemble the template sequences, are defined by the indicated "start" and "stop" nucleotide positions along each template.

Table 6 shows the tissue distribution profiles for the templates of the invention.

Table 7 shows the sequence identification numbers (SEQ ID NO:s) corresponding to the polypeptides of the present invention, along with the reading frames used to obtain the polypeptide  
20 segments, the lengths of the polypeptide segments, the "start" and "stop" nucleotide positions of the polynucleotide sequences used to define the encoded polypeptide segments, the GenBank hits (GI Numbers), probability scores, and functional annotations corresponding to the GenBank hits.

Table 8 summarizes the bioinformatics tools which are useful for analysis of the polynucleotides of the present invention. The first column of Table 8 lists analytical tools, programs,  
25 and algorithms, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the  
homology between two sequences).

30

## DETAILED DESCRIPTION OF THE INVENTION

Before the nucleic acid sequences and methods are presented, it is to be understood that this invention is not limited to the particular machines, methods, and materials described. Although particular embodiments are described, machines, methods, and materials similar or equivalent to these

embodiments may be used to practice the invention. The preferred machines, methods, and materials set forth are not intended to limit the scope of the invention which is limited only by the appended claims.

The singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. All technical and scientific terms have the meanings commonly understood by one of ordinary skill in the art. All publications are incorporated by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are presented and which might be used in connection with the invention. Nothing in the specification is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

### **Definitions**

As used herein, the lower case "mddt" refers to a nucleic acid sequence, while the upper case "MDDT" refers to an amino acid sequence encoded by mddt. A "full-length" mddt refers to a nucleic acid sequence containing the entire coding region of a gene endogenously expressed in human tissue.

"Adjuvants" are materials such as Freund's adjuvant, mineral gels (aluminum hydroxide), and surface active substances (lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol) which may be administered to increase a host's immunological response.

"Allele" refers to an alternative form of a nucleic acid sequence. Alleles result from a "mutation," a change or an alternative reading of the genetic code. Any given gene may have none, one, or many allelic forms. Mutations which give rise to alleles include deletions, additions, or substitutions of nucleotides. Each of these changes may occur alone, or in combination with the others, one or more times in a given nucleic acid sequence. The present invention encompasses allelic mddt.

"Amino acid sequence" refers to a peptide, a polypeptide, or a protein of either natural or synthetic origin. The amino acid sequence is not limited to the complete, endogenous amino acid sequence and may be a fragment, epitope, variant, or derivative of a protein expressed by a nucleic acid sequence.

"Amplification" refers to the production of additional copies of a sequence and is carried out using polymerase chain reaction (PCR) technologies well known in the art.

"Antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind MDDT polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or peptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and

can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

5 "Antisense sequence" refers to a sequence capable of specifically hybridizing to a target sequence. The antisense sequence may include DNA, RNA, or any nucleic acid mimic or analog such as peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine.

10 "Antisense sequence" refers to a sequence capable of specifically hybridizing to a target sequence. The antisense sequence can be DNA, RNA, or any nucleic acid mimic or analog.

"Antisense technology" refers to any technology which relies on the specific hybridization of an antisense sequence to a target sequence.

15 A "bin" is a portion of computer memory space used by a computer program for storage of data, and bounded in such a manner that data stored in a bin may be retrieved by the program.

"Biologically active" refers to an amino acid sequence having a structural, regulatory, or biochemical function of a naturally occurring amino acid sequence.

20 "Clone joining" is a process for combining gene bins based upon the bins' containing sequence information from the same clone. The sequences may assemble into a primary gene transcript as well as one or more splice variants.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing (5'-A-G-T-3' pairs with its complement 3'-T-C-A-5').

25 A "component sequence" is a nucleic acid sequence selected by a computer program such as PHRED and used to assemble a consensus or template sequence from one or more component sequences.

A "consensus sequence" or "template sequence" is a nucleic acid sequence which has been assembled from overlapping sequences, using a computer program for fragment assembly such as the GELVIEW fragment assembly system (Genetics Computer Group (GCG), Madison WI) or using a relational database management system (RDMS).

30 "Conservative amino acid substitutions" are those substitutions that, when made, least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative substitutions.

35

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
5	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
10	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
15	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
20	Val	Ile, Leu, Thr

Conservative substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain.

“Deletion” refers to a change in either a nucleic acid sequence in which at least one nucleotide or amino acid residue, respectively, is absent.

“Derivative” refers to the chemical modification of a nucleic acid sequence, such as by replacement of hydrogen by an alkyl, acyl, amino, hydroxyl, or other group.

“Differential expression” refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

The terms “element” and “array element” refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

“E-value” refers to the statistical probability that a match between two sequences occurred by chance.

“Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of mddt or MDDT which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 10 to 1000 contiguous amino acid residues or nucleotides. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous amino acid residues or nucleotides in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing and the figures, may be encompassed by the present embodiments.

A fragment of mddt comprises a region of unique polynucleotide sequence that specifically identifies mddt, for example, as distinct from any other sequence in the same genome. A fragment of mddt is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish mddt from related polynucleotide sequences. The precise length of a fragment of mddt and the region of mddt to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of MDDT is encoded by a fragment of mddt. A fragment of MDDT comprises a region of unique amino acid sequence that specifically identifies MDDT. For example, a fragment of MDDT is useful as an immunogenic peptide for the development of antibodies that specifically recognize MDDT. The precise length of a fragment of MDDT and the region of MDDT to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" nucleotide sequence is one containing at least a start site for translation to a protein sequence, followed by an open reading frame and a stop site, and encoding a "full length" polypeptide.

"Hit" refers to a sequence whose annotation will be used to describe a given template. Criteria for selecting the top hit are as follows: if the template has one or more exact nucleic acid matches, the top hit is the exact match with highest percent identity. If the template has no exact matches but has significant protein hits, the top hit is the protein hit with the lowest E-value. If the template has no significant protein hits, but does have significant non-exact nucleotide hits, the top hit is the nucleotide hit with the lowest E-value.

"Homology" refers to sequence similarity either between a reference nucleic acid sequence and at least a fragment of an mddt or between a reference amino acid sequence and a fragment of an MDDT.

“Hybridization” refers to the process by which a strand of nucleotides anneals with a complementary strand through base pairing. Specific hybridization is an indication that two nucleic acid sequences share a high degree of identity. Specific hybridization complexes form under defined annealing conditions, and remain hybridized after the “washing” step. The defined hybridization conditions include the annealing conditions and the washing step(s), the latter of which is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid probes that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency.

Generally, stringency of hybridization is expressed with reference to the temperature under which the wash step is carried out. Generally, such wash temperatures are selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization is well known and can be found in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, or 55°C may be used. SSC concentration may be varied from about 0.2 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, denatured salmon sperm DNA at about 100-200 µg/ml. Useful variations on these conditions will be readily apparent to those skilled in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their resultant proteins.

Other parameters, such as temperature, salt concentration, and detergent concentration may be varied to achieve the desired stringency. Denaturants, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as RNA:DNA hybridizations. Appropriate hybridization conditions are routinely determinable by one of ordinary skill in the art.

“Immunologically active” or “immunogenic” describes the potential for a natural, recombinant, or synthetic peptide, epitope, polypeptide, or protein to induce antibody production in appropriate animals, cells, or cell lines.

"Insertion" or "addition" refers to a change in either a nucleic or amino acid sequence in which at least one nucleotide or residue, respectively, is added to the sequence.

"Labeling" refers to the covalent or noncovalent joining of a polynucleotide, polypeptide, or antibody with a reporter molecule capable of producing a detectable or measurable signal.

5 "Microarray" is any arrangement of nucleic acids, amino acids, antibodies, etc., on a substrate. The substrate may be a solid support such as beads, glass, paper, nitrocellulose, nylon, or an appropriate membrane.

"Linkers" are short stretches of nucleotide sequence which may be added to a vector or an mddt to create restriction endonuclease sites to facilitate cloning. "Polylinkers" are engineered to  
10 incorporate multiple restriction enzyme sites and to provide for the use of enzymes which leave 5' or 3' overhangs (e.g., BamHI, EcoRI, and HindIII) and those which provide blunt ends (e.g., EcoRV, SnaBI, and StuI).

"Naturally occurring" refers to an endogenous polynucleotide or polypeptide that may be isolated from viruses or prokaryotic or eukaryotic cells.

15 "Nucleic acid sequence" refers to the specific order of nucleotides joined by phosphodiester bonds in a linear, polymeric arrangement. Depending on the number of nucleotides, the nucleic acid sequence can be considered an oligomer, oligonucleotide, or polynucleotide. The nucleic acid can be DNA, RNA, or any nucleic acid analog, such as PNA, may be of genomic or synthetic origin, may be either double-stranded or single-stranded, and can represent either the sense or antisense  
20 (complementary) strand.

"Oligomer" refers to a nucleic acid sequence of at least about 6 nucleotides and as many as about 60 nucleotides, preferably about 15 to 40 nucleotides, and most preferably between about 20 and 30 nucleotides, that may be used in hybridization or amplification technologies. Oligomers may be used as, e.g., primers for PCR, and are usually chemically synthesized.

25 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

30 "Peptide nucleic acid" (PNA) refers to a DNA mimic in which nucleotide bases are attached to a pseudopeptide backbone to increase stability. PNAs, also designated antigene agents, can prevent gene expression by targeting complementary messenger RNA.

The phrases "percent identity" and "% identity", as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a  
35 standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in



the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and Sharp, P.M. (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequence pairs.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to determine alignment between a known polynucleotide sequence and other sequences on a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2/>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*  
*Reward for match: 1*  
*Penalty for mismatch: -2*  
*Open Gap: 5 and Extension Gap: 2 penalties*  
*Gap x drop-off: 50*  
*Expect: 10*  
*Word Size: 11*  
*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at

least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in figures or Sequence Listings, may be used to describe a length over which percentage identity may be measured.

5           Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

10           The phrases "percent identity" and "% identity", as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the hydrophobicity and acidity of the substituted residue, thus preserving the structure (and therefore function) of the folded polypeptide.

15           Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default  
20           residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

          Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) with blastp set at default parameters. Such default parameters may be, for example:

25           *Matrix: BLOSUM62*  
          *Open Gap: 11 and Extension Gap: 1 penalty*  
          *Gap x drop-off: 50*  
          *Expect: 10*  
          *Word Size: 3*  
30           *Filter: on*

          Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least  
35           150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment

length supported by the sequences shown herein, in figures or Sequence Listings, may be used to describe a length over which percentage identity may be measured.

"Post-translational modification" of an MDDT may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu and the MDDT.

"Probe" refers to mddt or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the figures and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel et al., 1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer

selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

"Purified" refers to molecules, either polynucleotides or polypeptides that are isolated or separated from their natural environment and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other compounds with which they are naturally associated.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

"Regulatory element" refers to a nucleic acid sequence from nontranslated regions of a gene, and includes enhancers, promoters, introns, and 3' untranslated regions, which interact with host proteins to carry out or regulate transcription or translation.

"Reporter" molecules are chemical or biochemical moieties used for labeling a nucleic acid, an amino acid, or an antibody. They include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

5        "Sample" is used in its broadest sense. Samples may contain nucleic or amino acids, antibodies, or other materials, and may be derived from any source (e.g., bodily fluids including, but not limited to, saliva, blood, and urine; chromosome(s), organelles, or membranes isolated from a cell; genomic DNA, RNA, or cDNA in solution or bound to a substrate; and cleared cells or tissues or blots or imprints from such cells or tissues).

10        "Specific binding" or "specifically binding" refers to the interaction between a protein or peptide and its agonist, antibody, antagonist, or other binding partner. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing epitope A, or the presence of free unlabeled A, in a reaction  
15        containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

"Substitution" refers to the replacement of at least one nucleotide or amino acid by a different nucleotide or amino acid.

20        "Substrate" refers to any suitable rigid or semi-rigid support including, e.g., membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles or capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular tissue or cell type under given conditions at a given time.

25        "Transformation" refers to a process by which exogenous DNA enters a recipient cell. Transformation may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the host cell being transformed.

30        "Transformants" include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as cells which transiently express inserted DNA or RNA.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid  
35        introduced by way of human intervention, such as by transgenic techniques well known in the art. The

nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, and plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 25% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 30%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. The variant may result in "conservative" amino acid changes which do not affect structural and/or chemical properties. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

In an alternative, variants of the polynucleotides of the present invention may be generated through recombinant methods. One possible method is a DNA shuffling technique such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of MDDT, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is

produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

In a particular embodiment, cDNA sequences derived from human tissues and cell lines were aligned based on nucleotide sequence identity and assembled into "consensus" or "template" sequences which are designated by the template identification numbers (template IDs) in column 2 of Table 2. The sequence identification numbers (SEQ ID NO:s) corresponding to the template IDs are shown in column 1. The template sequences have similarity to GenBank sequences, or "hits," as designated by the GI Numbers in column 3. The statistical probability of each GenBank hit is indicated by a probability score in column 4, and the functional annotation corresponding to each GenBank hit is listed in column 5.

The invention incorporates the nucleic acid sequences of these templates as disclosed in the Sequence Listing and the use of these sequences in the diagnosis and treatment of disease states characterized by defects in disease detection and treatment molecules. The invention further utilizes these sequences in hybridization and amplification technologies, and in particular, in technologies which assess gene expression patterns correlated with specific cells or tissues and their responses *in vivo* or *in vitro* to pharmaceutical agents, toxins, and other treatments. In this manner, the sequences of the present invention are used to develop a transcript image for a particular cell or tissue.

### Derivation of Nucleic Acid Sequences

cDNA was isolated from libraries constructed using RNA derived from normal and diseased human tissues and cell lines. The human tissues and cell lines used for cDNA library construction were selected from a broad range of sources to provide a diverse population of cDNAs representative of gene transcription throughout the human body. Descriptions of the human tissues and cell lines used for cDNA library construction are provided in the LIFESEQ database (Incyte Genomics, Inc. (Incyte), Palo Alto CA). Human tissues were broadly selected from, for example, cardiovascular, dermatologic, endocrine, gastrointestinal, hematopoietic/immune system, musculoskeletal, neural, reproductive, and urologic sources.

Cell lines used for cDNA library construction were derived from, for example, leukemic cells, teratocarcinomas, neuroepitheliomas, cervical carcinoma, lung fibroblasts, and endothelial cells. Such cell lines include, for example, THP-1, Jurkat, HUVEC, hNT2, WI38, HeLa, and other cell lines commonly used and available from public depositories (American Type Culture Collection, Manassas VA). Prior to mRNA isolation, cell lines were untreated, treated with a pharmaceutical agent such as 5'-aza-2'-deoxycytidine, treated with an activating agent such as lipopolysaccharide in the case of leukocytic cell lines, or, in the case of endothelial cell lines, subjected to shear stress.

#### Sequencing of the cDNAs

Methods for DNA sequencing are well known in the art. Conventional enzymatic methods employ the Klenow fragment of DNA polymerase I, SEQUENASE DNA polymerase (U.S.

Biochemical Corporation, Cleveland OH), Taq polymerase (Applied Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Inc. (Amersham Pharmacia Biotech), Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies Inc. (Life Technologies), Gaithersburg MD), to extend the nucleic acid sequence from an oligonucleotide primer annealed to the DNA

template of interest. Methods have been developed for the use of both single-stranded and double-stranded templates. Chain termination reaction products may be electrophoresed on urea-polyacrylamide gels and detected either by autoradiography (for radioisotope-labeled nucleotides) or by fluorescence (for fluorophore-labeled nucleotides). Automated methods for mechanized reaction preparation, sequencing, and analysis using fluorescence detection methods have been developed.

Machines used to prepare cDNAs for sequencing can include the MICROLAB 2200 liquid transfer system (Hamilton Company (Hamilton), Reno NV), Peltier thermal cycler (PTC200; MJ Research, Inc. (MJ Research), Watertown MA), and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing can be carried out using, for example, the ABI 373 or 377 (Applied Biosystems) or MEGABACE 1000 (Molecular Dynamics, Inc. (Molecular Dynamics), Sunnyvale CA) DNA sequencing systems, or other automated and manual sequencing systems well known in the



art.

The nucleotide sequences of the Sequence Listing have been prepared by current, state-of-the-art, automated methods and, as such, may contain occasional sequencing errors or unidentified nucleotides. Such unidentified nucleotides are designated by an N. These infrequent unidentified bases do not represent a hindrance to practicing the invention for those skilled in the art. Several methods employing standard recombinant techniques may be used to correct errors and complete the missing sequence information. (See, e.g., those described in Ausubel, F.M. et al. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY; and Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY.)

#### Assembly of cDNA Sequences

Human polynucleotide sequences may be assembled using programs or algorithms well known in the art. Sequences to be assembled are related, wholly or in part, and may be derived from a single or many different transcripts. Assembly of the sequences can be performed using such programs as PHRAP (Phils Revised Assembly Program) and the GELVIEW fragment assembly system (GCG), or other methods known in the art.

Alternatively, cDNA sequences are used as "component" sequences that are assembled into "template" or "consensus" sequences as follows. Sequence chromatograms are processed, verified, and quality scores are obtained using PHRED. Raw sequences are edited using an editing pathway known as Block 1 (See, e.g., the LIFESEQ Assembled User Guide, Incyte Genomics, Palo Alto, CA). A series of BLAST comparisons is performed and low-information segments and repetitive elements (e.g., dinucleotide repeats, Alu repeats, etc.) are replaced by "n's", or masked, to prevent spurious matches. Mitochondrial and ribosomal RNA sequences are also removed. The processed sequences are then loaded into a relational database management system (RDMS) which assigns edited sequences to existing templates, if available. When additional sequences are added into the RDMS, a process is initiated which modifies existing templates or creates new templates from works in progress (i.e., nonfinal assembled sequences) containing queued sequences or the sequences themselves. After the new sequences have been assigned to templates, the templates can be merged into bins. If multiple templates exist in one bin, the bin can be split and the templates reannotated.

Once gene bins have been generated based upon sequence alignments, bins are "clone joined" based upon clone information. Clone joining occurs when the 5' sequence of one clone is present in one bin and the 3' sequence from the same clone is present in a different bin, indicating that the two bins should be merged into a single bin. Only bins which share at least two different clones are merged.

A resultant template sequence may contain either a partial or a full length open reading frame,

or all or part of a genetic regulatory element. This variation is due in part to the fact that the full length cDNAs of many genes are several hundred, and sometimes several thousand, bases in length. With current technology, cDNAs comprising the coding regions of large genes cannot be cloned because of vector limitations, incomplete reverse transcription of the mRNA, or incomplete "second strand" synthesis. Template sequences may be extended to include additional contiguous sequences derived from the parent RNA transcript using a variety of methods known to those of skill in the art. Extension may thus be used to achieve the full length coding sequence of a gene.

#### Analysis of the cDNA Sequences

The cDNA sequences are analyzed using a variety of programs and algorithms which are well known in the art. (See, e.g., Ausubel, 1997, supra, Chapter 7.7; Meyers, R.A. (Ed.) (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853; and Table 8.) These analyses comprise both reading frame determinations, e.g., based on triplet codon periodicity for particular organisms (Fickett, J.W. (1982) *Nucleic Acids Res.* 10:5303-5318); analyses of potential start and stop codons; and homology searches.

Computer programs known to those of skill in the art for performing computer-assisted searches for amino acid and nucleic acid sequence similarity, include, for example, Basic Local Alignment Search Tool (BLAST; Altschul, S.F. (1993) *J. Mol. Evol.* 36:290-300; Altschul, S.F. et al. (1990) *J. Mol. Biol.* 215:403-410). BLAST is especially useful in determining exact matches and comparing two sequence fragments of arbitrary but equal lengths, whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user (Karin, S. et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:841-845). Using an appropriate search tool (e.g., BLAST or HMM), GenBank, SwissProt, BLOCKS, PFAM and other databases may be searched for sequences containing regions of homology to a query mddt or MDDT of the present invention.

Other approaches to the identification, assembly, storage, and display of nucleotide and polypeptide sequences are provided in "Relational Database for Storing Biomolecule Information," U.S.S.N. 08/947,845, filed October 9, 1997; "Project-Based Full-Length Biomolecular Sequence Database," U.S.S.N. 08/811,758, filed March 6, 1997; and "Relational Database and System for Storing Information Relating to Biomolecular Sequences," U.S.S.N. 09/034,807, filed March 4, 1998, all of which are incorporated by reference herein in their entirety.

Protein hierarchies can be assigned to the putative encoded polypeptide based on, e.g., motif, BLAST, or biological analysis. Methods for assigning these hierarchies are described, for example, in "Database System Employing Protein Function Hierarchies for Viewing Biomolecular Sequence Data," U.S.S.N. 08/812,290, filed March 6, 1997, incorporated herein by reference.

### Human Disease Detection and Treatment Molecule Sequences

The mddt of the present invention may be used for a variety of diagnostic and therapeutic purposes. For example, an mddt may be used to diagnose a particular condition, disease, or disorder associated with disease detection and treatment molecules. Such conditions, diseases, and disorders include, but are not limited to, a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an autoimmune/inflammatory disorder, such as actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma. The mddt can be used to detect the presence of, or to quantify the amount of, an mddt-related polynucleotide in a sample. This information is then compared to information obtained from appropriate reference samples, and a diagnosis is established. Alternatively, a polynucleotide complementary to a given mddt can inhibit or inactivate a therapeutically relevant gene related to the mddt.

### Analysis of mddt Expression Patterns

The expression of mddt may be routinely assessed by hybridization-based methods to determine, for example, the tissue-specificity, disease-specificity, or developmental stage-specificity of

mddt expression. For example, the level of expression of mddt may be compared among different cell types or tissues, among diseased and normal cell types or tissues, among cell types or tissues at different developmental stages, or among cell types or tissues undergoing various treatments. This type of analysis is useful, for example, to assess the relative levels of mddt expression in fully or partially differentiated cells or tissues, to determine if changes in mddt expression levels are correlated with the development or progression of specific disease states, and to assess the response of a cell or tissue to a specific therapy, for example, in pharmacological or toxicological studies. Methods for the analysis of mddt expression are based on hybridization and amplification technologies and include membrane-based procedures such as northern blot analysis, high-throughput procedures that utilize, for example, microarrays, and PCR-based procedures.

#### Hybridization and Genetic Analysis

The mddt, their fragments, or complementary sequences, may be used to identify the presence of and/or to determine the degree of similarity between two (or more) nucleic acid sequences. The mddt may be hybridized to naturally occurring or recombinant nucleic acid sequences under appropriately selected temperatures and salt concentrations. Hybridization with a probe based on the nucleic acid sequence of at least one of the mddt allows for the detection of nucleic acid sequences, including genomic sequences, which are identical or related to the mddt of the Sequence Listing. Probes may be selected from non-conserved or unique regions of at least one of the polynucleotides of SEQ ID NO:1-252 and tested for their ability to identify or amplify the target nucleic acid sequence using standard protocols.

Polynucleotide sequences that are capable of hybridizing, in particular, to those shown in SEQ ID NO:1-252 and fragments thereof, can be identified using various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions are discussed in "Definitions."

A probe for use in Southern or northern hybridization may be derived from a fragment of an mddt sequence, or its complement, that is up to several hundred nucleotides in length and is either single-stranded or double-stranded. Such probes may be hybridized in solution to biological materials such as plasmids, bacterial, yeast, or human artificial chromosomes, cleared or sectioned tissues, or to artificial substrates containing mddt. Microarrays are particularly suitable for identifying the presence of and detecting the level of expression for multiple genes of interest by examining gene expression correlated with, e.g., various stages of development, treatment with a drug or compound, or disease progression. An array analogous to a dot or slot blot may be used to arrange and link polynucleotides to the surface of a substrate using one or more of the following: mechanical (vacuum), chemical, thermal, or UV bonding procedures. Such an array may contain any number of mddt and may be

produced by hand or by using available devices, materials, and machines.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

Probes may be labeled by either PCR or enzymatic techniques using a variety of commercially available reporter molecules. For example, commercial kits are available for radioactive and chemiluminescent labeling (Amersham Pharmacia Biotech) and for alkaline phosphatase labeling (Life Technologies). Alternatively, mddt may be cloned into commercially available vectors for the production of RNA probes. Such probes may be transcribed in the presence of at least one labeled nucleotide (e.g., <sup>32</sup>P-ATP, Amersham Pharmacia Biotech).

Additionally the polynucleotides of SEQ ID NO:1-252 or suitable fragments thereof can be used to isolate full length cDNA sequences utilizing hybridization and/or amplification procedures well known in the art, e.g., cDNA library screening, PCR amplification, etc. The molecular cloning of such full length cDNA sequences may employ the method of cDNA library screening with probes using the hybridization, stringency, washing, and probing strategies described above and in Ausubel, supra, Chapters 3, 5, and 6. These procedures may also be employed with genomic libraries to isolate genomic sequences of mddt in order to analyze, e.g., regulatory elements.

#### Genetic Mapping

Gene identification and mapping are important in the investigation and treatment of almost all conditions, diseases, and disorders. Cancer, cardiovascular disease, Alzheimer's disease, arthritis, diabetes, and mental illnesses are of particular interest. Each of these conditions is more complex than the single gene defects of sickle cell anemia or cystic fibrosis, with select groups of genes being predictive of predisposition for a particular condition, disease, or disorder. For example, cardiovascular disease may result from malfunctioning receptor molecules that fail to clear cholesterol from the bloodstream, and diabetes may result when a particular individual's immune system is activated by an infection and attacks the insulin-producing cells of the pancreas. In some studies, Alzheimer's disease has been linked to a gene on chromosome 21; other studies predict a different gene and location. Mapping of disease genes is a complex and reiterative process and generally proceeds from genetic linkage analysis to physical mapping.

As a condition is noted among members of a family, a genetic linkage map traces parts of chromosomes that are inherited in the same pattern as the condition. Statistics link the inheritance of particular conditions to particular regions of chromosomes, as defined by RFLP or other markers.

(See, for example, Lander, E. S. and Botstein, D. (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.) Occasionally, genetic markers and their locations are known from previous studies. More often, however, the markers are simply stretches of DNA that differ among individuals. Examples of genetic linkage maps can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site.

In another embodiment of the invention, mddt sequences may be used to generate hybridization probes useful in chromosomal mapping of naturally occurring genomic sequences. Either coding or noncoding sequences of mddt may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of an mddt coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Meyers, supra, pp. 965-968.) Correlation between the location of mddt on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The mddt sequences may also be used to detect polymorphisms that are genetically linked to the inheritance of a particular condition, disease, or disorder.

In situ hybridization of chromosomal preparations and genetic mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending existing genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of the corresponding human chromosome is not known. These new marker sequences can be mapped to human chromosomes and may provide valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once a disease or syndrome has been crudely correlated by genetic linkage with a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequences of the subject invention may also be used to detect differences in chromosomal architecture due to translocation, inversion, etc., among normal, carrier, or affected individuals.

Once a disease-associated gene is mapped to a chromosomal region, the gene must be cloned

in order to identify mutations or other alterations (e.g., translocations or inversions) that may be correlated with disease. This process requires a physical map of the chromosomal region containing the disease-gene of interest along with associated markers. A physical map is necessary for determining the nucleotide sequence of and order of marker genes on a particular chromosomal region. Physical mapping techniques are well known in the art and require the generation of overlapping sets of cloned DNA fragments from a particular organelle, chromosome, or genome. These clones are analyzed to reconstruct and catalog their order. Once the position of a marker is determined, the DNA from that region is obtained by consulting the catalog and selecting clones from that region. The gene of interest is located through positional cloning techniques using hybridization or similar methods.

#### Diagnostic Uses

The mddt of the present invention may be used to design probes useful in diagnostic assays. Such assays, well known to those skilled in the art, may be used to detect or confirm conditions, disorders, or diseases associated with abnormal levels of mddt expression. Labeled probes developed from mddt sequences are added to a sample under hybridizing conditions of desired stringency. In some instances, mddt, or fragments or oligonucleotides derived from mddt, may be used as primers in amplification steps prior to hybridization. The amount of hybridization complex formed is quantified and compared with standards for that cell or tissue. If mddt expression varies significantly from the standard, the assay indicates the presence of the condition, disorder, or disease. Qualitative or quantitative diagnostic methods may include northern, dot blot, or other membrane or dip-stick based technologies or multiple-sample format technologies such as PCR, enzyme-linked immunosorbent assay (ELISA)-like, pin, or chip-based assays.

The probes described above may also be used to monitor the progress of conditions, disorders, or diseases associated with abnormal levels of mddt expression, or to evaluate the efficacy of a particular therapeutic treatment. The candidate probe may be identified from the mddt that are specific to a given human tissue and have not been observed in GenBank or other genome databases. Such a probe may be used in animal studies, preclinical tests, clinical trials, or in monitoring the treatment of an individual patient. In a typical process, standard expression is established by methods well known in the art for use as a basis of comparison, samples from patients affected by the disorder or disease are combined with the probe to evaluate any deviation from the standard profile, and a therapeutic agent is administered and effects are monitored to generate a treatment profile. Efficacy is evaluated by determining whether the expression progresses toward or returns to the standard normal pattern. Treatment profiles may be generated over a period of several days or several months. Statistical methods well known to those skilled in the art may be used to determine the significance of

such therapeutic agents.

The polynucleotides are also useful for identifying individuals from minute biological samples, for example, by matching the RFLP pattern of a sample's DNA to that of an individual's DNA. The polynucleotides of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, an individual can be identified through a unique set of DNA sequences. Once a unique ID database is established for an individual, positive identification of that individual can be made from extremely small tissue samples.

In a particular aspect, oligonucleotide primers derived from the mddt of the invention may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from mddt are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequences of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

DNA-based identification techniques are critical in forensic technology. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using, e.g., PCR, to identify individuals. (See, e.g., Erlich, H. (1992) PCR Technology, Freeman and Co., New York, NY). Similarly, polynucleotides of the present invention can be used as polymorphic markers.

There is also a need for reagents capable of identifying the source of a particular tissue. Appropriate reagents can comprise, for example, DNA probes or primers prepared from the sequences of the present invention that are specific for particular tissues. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to



screen tissue cultures for contamination.

The polynucleotides of the present invention can also be used as molecular weight markers on nucleic acid gels or Southern blots, as diagnostic probes for the presence of a specific mRNA in a particular cell type, in the creation of subtracted cDNA libraries which aid in the discovery of novel polynucleotides, in selection and synthesis of oligomers for attachment to an array or other support, and as an antigen to elicit an immune response.

#### Disease Model Systems Using mddt

The mddt of the invention or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292).

The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

The mddt of the invention may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

The mddt of the invention can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of mddt is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress mddt, resulting, e.g., in the secretion of MDDT in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

### Screening Assays

5 MDDT encoded by polynucleotides of the present invention may be used to screen for molecules that bind to or are bound by the encoded polypeptides. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the bound molecule. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

10 Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a ligand or fragment thereof, a natural substrate, or a structural or functional mimetic. (See, Coligan et al., (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or to at least a fragment of the receptor, e.g., the active site. In either case, the molecule can be rationally designed using known techniques. Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing the polypeptide or cell membrane fractions  
15 which contain the expressed polypeptide are then contacted with a test compound and binding, stimulation, or inhibition of activity of either the polypeptide or the molecule is analyzed.

An assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. Alternatively, the assay may assess binding in the presence of a labeled competitor.

20 Additionally, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

25 Preferably, an ELISA assay using, e.g., a monoclonal or polyclonal antibody, can measure polypeptide level in a sample. The antibody can measure polypeptide level by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

30 All of the above assays can be used in a diagnostic or prognostic context. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

### Expression

5 In order to express a biologically active MDDT, the nucleotide sequences encoding MDDT or fragments thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding MDDT and appropriate transcriptional and  
10 translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, supra, Chapters 4, 8, 16, and 17; and Ausubel, supra, Chapters 9, 10, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding MDDT. These include, but are not limited to, microorganisms such as bacteria transformed  
15 with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal (mammalian) cell systems. (See, e.g., Sambrook, supra; Ausubel, 1995, supra, Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105;  
20 The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al., (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The  
25 invention is not limited by the host cell employed.

For long term production of recombinant proteins in mammalian systems, stable expression of  
35 MDDT in cell lines is preferred. For example, sequences encoding MDDT can be transformed into

cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Any number of selection systems may be used to recover transformed cell lines. (See, e.g., Wigler, M. et al. (1977) *Cell* 11:223-232; Lowy, I. et al. (1980) *Cell* 22:817-823.; Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci. USA* 77:3567-3570; Colbere-Garapin, F. et al. (1981) *J. Mol. Biol.* 150:1-14; Hartman, S.C. and R.C.Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:8047-8051; Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

#### Therapeutic Uses of mddt

The mddt of the invention may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and Somia, N. (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in mddt expression or regulation causes disease, the expression of mddt from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in mddt are treated by constructing mammalian expression vectors comprising mddt and introducing these vectors by mechanical means into mddt-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and Anderson, W.F. (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and Récipon, H. (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of mddt include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). The mddt of the invention  
5 may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and Bujard, H. (1992) Proc. Natl. Acad. Sci. U.S.A. 89:5547-5551; Gossen, M. et al., (1995) Science 268:1766-1769; Rossi, F.M.V. and Blau, H.M. (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid  
10 (Invitrogen); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding MDDT from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID  
15 TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and Eb, A.J. (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these  
20 standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to mddt expression are treated by constructing a retrovirus vector consisting of (i) mddt under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional  
25 retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. U.S.A. 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous  
30 envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and Miller, A.D. (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging  
35 cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of

a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. U.S.A. 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver mddt to cells which have one or more genetic abnormalities with respect to the expression of mddt. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and Somia, N. (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver mddt to target cells which have one or more genetic abnormalities with respect to the expression of mddt. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing mddt to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W. F. et al. 1999 J. Virol. 73:519-532 and Xu, H. et al., (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver mddt to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H.

and Li, K-J. (1998) *Curr. Opin. Biotech.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase).

5 Similarly, inserting mddt into the alphavirus genome in place of the capsid-coding region results in the production of a large number of mddt RNAs and the synthesis of high levels of MDDT in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of

10 the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of mddt into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the

15 art.

#### Antibodies

Anti-MDDT antibodies may be used to analyze protein expression levels. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, and Fab fragments. For

20 descriptions of and protocols of antibody technologies, see, e.g., Pound J.D. (1998) Immunochemical Protocols, Humana Press, Totowa, NJ.

The amino acid sequence encoded by the mddt of the Sequence Listing may be analyzed by appropriate software (e.g., LASERGENE NAVIGATOR software, DNASTAR) to determine regions of high immunogenicity. The optimal sequences for immunization are selected from the C-

25 terminus, the N-terminus, and those intervening, hydrophilic regions of the polypeptide which are likely to be exposed to the external environment when the polypeptide is in its natural conformation. Analysis used to select appropriate epitopes is also described by Ausubel (1997, supra, Chapter 11.7). Peptides used for antibody induction do not need to have biological activity; however, they must be antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of

30 at least five amino acids, preferably at least 10 amino acids, and most preferably at least 15 amino acids. A peptide which mimics an antigenic fragment of the natural polypeptide may be fused with another protein such as keyhole hemolimpet cyanin (KLH; Sigma, St. Louis MO) for antibody production. A peptide encompassing an antigenic region may be expressed from an mddt, synthesized as described above, or purified from human cells.

35 Procedures well known in the art may be used for the production of antibodies. Various hosts

including mice, goats, and rabbits, may be immunized by injection with a peptide. Depending on the host species, various adjuvants may be used to increase immunological response.

In one procedure, peptides about 15 residues in length may be synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using fmoc-chemistry and coupled to KLH (Sigma) by  
5 reaction with M-maleimidobenzoyl-N-hydroxysuccinimide ester (Ausubel, 1995, supra). Rabbits are immunized with the peptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity by binding the peptide to plastic, blocking with 1% bovine serum albumin (BSA), reacting with rabbit antisera, washing, and reacting with radioiodinated goat anti-rabbit IgG. Antisera with antipeptide activity are tested for anti-MDDT activity using protocols well known in the  
10 art, including ELISA, radioimmunoassay (RIA), and immunoblotting.

In another procedure, isolated and purified peptide may be used to immunize mice (about 100 µg of peptide) or rabbits (about 1 mg of peptide). Subsequently, the peptide is radioiodinated and used to screen the immunized animals' B-lymphocytes for production of antipeptide antibodies. Positive cells are then used to produce hybridomas using standard techniques. About 20 mg of peptide is  
15 sufficient for labeling and screening several thousand clones. Hybridomas of interest are detected by screening with radioiodinated peptide to identify those fusions producing peptide-specific monoclonal antibody. In a typical protocol, wells of a multi-well plate (FAST, Becton-Dickinson, Palo Alto, CA) are coated with affinity-purified, specific rabbit-anti-mouse (or suitable anti-species IgG) antibodies at 10 mg/ml. The coated wells are blocked with 1% BSA and washed and exposed to supernatants from  
20 hybridomas. After incubation, the wells are exposed to radiolabeled peptide at 1 mg/ml.

Clones producing antibodies bind a quantity of labeled peptide that is detectable above background. Such clones are expanded and subjected to 2 cycles of cloning. Cloned hybridomas are injected into pristane-treated mice to produce ascites, and monoclonal antibody is purified from the ascitic fluid by affinity chromatography on protein A (Amersham Pharmacia Biotech). Several  
25 procedures for the production of monoclonal antibodies, including in vitro production, are described in Pound (supra). Monoclonal antibodies with antipeptide activity are tested for anti-MDDT activity using protocols well known in the art, including ELISA, RIA, and immunoblotting.

Antibody fragments containing specific binding sites for an epitope may also be generated. For example, such fragments include, but are not limited to, the F(ab')<sub>2</sub> fragments produced by pepsin  
30 digestion of the antibody molecule, and the Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, construction of Fab expression libraries in filamentous bacteriophage allows rapid and easy identification of monoclonal fragments with desired specificity (Pound, supra, Chaps. 45-47). Antibodies generated against polypeptide encoded by mddt can be used to purify and characterize full-length MDDT protein and its activity, binding partners, etc.

35



### Assays Using Antibodies

Anti-MDDT antibodies may be used in assays to quantify the amount of MDDT found in a particular human cell. Such assays include methods utilizing the antibody and a label to detect expression level under normal or disease conditions. The peptides and antibodies of the invention may be used with or without modification or labeled by joining them, either covalently or noncovalently, with a reporter molecule.

Protocols for detecting and measuring protein expression using either polyclonal or monoclonal antibodies are well known in the art. Examples include ELISA, RIA, and fluorescent activated cell sorting (FACS). Such immunoassays typically involve the formation of complexes between the MDDT and its specific antibody and the measurement of such complexes. These and other assays are described in Pound (supra).

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Serial No. 60/230,517, U.S. Serial No. 60/230,599, U.S. Serial No. 60/230,514, U.S. Serial No. 60/231,167, U.S. Serial No. 60/230,598, U.S. Serial No. 60/230,988, U.S. Serial No. 60/230,518, U.S. Serial No. 60/230,515, U.S. Serial No. 60/229,751, U.S. Serial No. 60/230,610, U.S. Serial No. 60/229,749, U.S. Serial No. 60/229,750, U.S. Serial No. 60/230,597, U.S. Serial No. 60/230,505, U.S. Serial No. 60/231,163, U.S. Serial No. 60/229,747, U.S. Serial No. 60/229,748, U.S. Serial No. 60/230,583, U.S. Serial No. 60/230,519, U.S. Serial No. 60/230,595, U.S. Serial No. 60/230,865, U.S. Serial No. 60/230,989, and U.S. Serial No. 60/230,951, are hereby expressly incorporated by reference.

## **EXAMPLES**

### **I. Construction of cDNA Libraries**

RNA was purchased from CLONTECH Laboratories, Inc. (Palo Alto CA) or isolated from various tissues. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In most cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated

TABLE 1

SEQ ID NO:	Template ID	SEQ ID NO:	ORF ID
160	U:1075296.1:2000SEP08	413	U:1075296.1.orf1:2000SEP08
161	U:1085501.1:2000SEP08	414	U:1085501.1.orf3:2000SEP08
162	U:1086181.1:2000SEP08	415	U:1086181.1.orf2:2000SEP08
163	U:1164493.1:2000SEP08	416	U:1164493.1.orf1:2000SEP08
164	U:1175097.1:2000SEP08	417	U:1175097.1.orf1:2000SEP08
165	U:1092948.1:2000SEP08	418	U:1092948.1.orf2:2000SEP08
166	U:380378.2:2000SEP08	419	U:380378.2.orf3:2000SEP08
167	U:1029674.1:2000SEP08	420	U:1029674.1.orf3:2000SEP08
168	U:2048601.3:2000SEP08	421	U:2048601.3.orf2:2000SEP08
169	U:1186208.1:2000SEP08	422	U:1186208.1.orf2:2000SEP08
170	U:1170753.1:2000SEP08	423	U:1170753.1.orf1:2000SEP08
171	U:1180908.1:2000SEP08	424	U:1180908.1.orf2:2000SEP08
172	U:1182900.2:2000SEP08	425	U:1182900.2.orf1:2000SEP08
173	U:1169548.2:2000SEP08	426	U:1169548.2.orf3:2000SEP08
174	U:1039974.1:2000SEP08	427	U:1039974.1.orf1:2000SEP08
175	U:1175765.2:2000SEP08	428	U:1175765.2.orf3:2000SEP08
176	U:313948.1:2000SEP08	429	U:313948.1.orf3:2000SEP08
177	U:335923.2:2000SEP08	430	U:335923.2.orf2:2000SEP08
178	U:345884.1:2000SEP08	431	U:345884.1.orf1:2000SEP08
179	U:417127.1:2000SEP08	432	U:417127.1.orf1:2000SEP08
180	U:451710.1:2000SEP08	433	U:451710.1.orf1:2000SEP08
181	U:406882.2:2000SEP08	434	U:406882.2.orf1:2000SEP08
182	U:728223.1:2000SEP08	435	U:728223.1.orf2:2000SEP08
183	U:289783.19:2000SEP08	436	U:289783.19.orf2:2000SEP08
184	U:235255.8:2000SEP08	437	U:235255.8.orf2:2000SEP08
185	U:237693.5:2000SEP08	438	U:237693.5.orf3:2000SEP08
186	U:433670.3:2000SEP08	439	U:433670.3.orf2:2000SEP08
187	U:202943.4:2000SEP08	440	U:202943.4.orf2:2000SEP08
188	U:068682.1:2000SEP08	441	U:068682.1.orf1:2000SEP08
189	U:203301.3:2000SEP08	442	U:203301.3.orf2:2000SEP08
190	U:020726.3:2000SEP08	443	U:020726.3.orf2:2000SEP08
191	U:027209.1:2000SEP08	444	U:027209.1.orf1:2000SEP08

		TABLE 2		Probability Score	Annotation
SEQ ID NO:	Template ID	GI Number			
177	U:335923.2:2000SEP08	g11990770		3.00E-73	(fl)(Homo sapiens) bA534G20.1.1 (novel protein similar to Lysozyme C-1 (1,4-beta-N- acylmuramidase C, Homo sapiens immunoglobulin receptor translocation associated protein 2c (IRTA2) mRNA, Homo sapiens, clone IMAGE:3352566, (fl)(Mercurialis annua) ribosomal Homo sapiens cDNA FLJ20081 fls, clone COL03242. (fl)(Oryza sativa) Homo sapiens, Similar to glucose regulated protein, 58 kDa, clone Homo sapiens clone IMAGE:72154 tRNA-guanine transglycosylase (TG) mRNA, (fl)(Homo sapiens) unnamed protein Homo sapiens mRNA for KIAA1479 protein, partial cds. Homo sapiens serologically defined breast cancer antigen NY- BR-38 mRNA, Homo sapiens serine/threonine kinase FKSG82 (FKSG82) mRNA, Homo sapiens mRNA for KIAA1548 protein, partial cds. 0 (Mus musculus)
178	U:345884.1:2000SEP08	g13591713		0	
179	U:417127.1:2000SEP08	g12652726		1.00E-66	
180	U:451710.1:2000SEP08	g13899057		1.00E-61	
181	U:406882.2:2000SEP08	g7019948		2.00E-57	
182	U:728223.1:2000SEP08	g2624328		2.00E-44	
183	U:289783.19:2000SEP08	g12654714		0	
184	U:235255.8:2000SEP08	g12597311		0	
185	U:237693.5:2000SEP08	g12406772		4.00E-52	
186	U:433670.3:2000SEP08	g14133250		0	
187	U:202943.4:2000SEP08	g12060829		0	
188	U:068682.1:2000SEP08	g13540325		0	
189	U:203301.3:2000SEP08	g10047160		0	
190	U:020726.3:2000SEP08	g12834087		1.00E-157	

TABLE 3

SEQ ID NO:	Template ID	Start	Stop	Frame	Pfam Hit	Pfam Description	E-value
183	LI:289783.19:2000SEP08	821	1003	forward 2	thioredoxin	Thioredoxin	1.70E-20
183	LI:289783.19:2000SEP08	747	800	forward 3	thioredoxin	Thioredoxin	5.10E-05
184	LI:235255.8:2000SEP08	419	919	forward 2	TGT	Queuine tRNA-ribosyltransferase	2.10E-24
184	LI:235255.8:2000SEP08	747	1346	forward 3	TGT	Queuine tRNA-ribosyltransferase	7.00E-15
185	LI:237693.5:2000SEP08	48	518	forward 3	abhydrolase_2	Phospholipase/Carboxylesterase	6.20E-07
186	LI:433670.3:2000SEP08	404	682	forward 2	Sema	Sema domain	3.80E-26
187	LI:202943.4:2000SEP08	119	223	forward 2	EGF	EGF-like domain	1.60E-05
187	LI:202943.4:2000SEP08	1304	1465	forward 2	sushi	Sushi domain (SCR repeat)	3.80E-18
188	LI:068682.1:2000SEP08	169	876	forward 1	pkinase	Protein kinase domain	1.70E-65
189	LI:203301.3:2000SEP08	570	716	forward 3	Band_41	FERM domain (Band 4.1 family)	1.50E-20
189	LI:203301.3:2000SEP08	365	577	forward 2	Band_41	FERM domain (Band 4.1 family)	6.30E-16
190	LI:020726.3:2000SEP08	542	1915	forward 2	MCT	Monocarboxylate transporter	1.80E-98
191	LI:027209.1:2000SEP08	379	987	forward 1	fibrinogen_C	Fibrinogen beta and gamma chains, C-terminal globular domain	7.80E-43
192	LI:108819.1:2000SEP08	316	828	forward 1	vwa	von Willebrand factor type A domain	1.90E-52
193	LI:021759.1:2000SEP08	1136	1246	forward 2	WD40	WD domain, G-beta repeat	1.70E-07
194	LI:1165967.1:2000SEP08	322	462	forward 1	Ribosomal_S27	Ribosomal protein S27a	1.50E-30
194	LI:1165967.1:2000SEP08	22	243	forward 1	ubiquitin	Ubiquitin family	4.50E-42
195	LI:1166315.1:2000SEP08	131	283	forward 2	pro_isomerase	Cyclophilin type peptidyl-prolyl cis-trans isomerase	1.10E-26
195	LI:1166315.1:2000SEP08	280	423	forward 1	pro_isomerase	Cyclophilin type peptidyl-prolyl cis-trans isomerase	5.50E-17
195	LI:1166315.1:2000SEP08	423	482	forward 3	pro_isomerase	Cyclophilin type peptidyl-prolyl cis-trans isomerase	4.10E-06
196	LI:204626.1:2000SEP08	322	1212	forward 1	Syntaxin	Syntaxin	8.60E-44
197	LI:801140.1:2000SEP08	516	584	forward 3	zf-C2H2	Zinc finger, C2H2 type	2.50E-08
198	LI:286639.1:2000SEP08	1295	1714	forward 2	actin	Actin	9.10E-64
198	LI:286639.1:2000SEP08	772	1296	forward 1	actin	Actin	1.60E-44
198	LI:286639.1:2000SEP08	630	755	forward 3	actin	Actin	1.10E-09
199	LI:288905.4:2000SEP08	285	602	forward 3	CH	Calponin homology (CH) domain	5.30E-27
200	LI:332161.1:2000SEP08	75	710	forward 3	ras	Ras family	2.00E-48
201	LI:184867.1:2000SEP08	365	478	forward 2	ubiquitin	Ubiquitin family	1.40E-04

TABLE 4

SEQ ID NO:	Template ID	Start	Stop	Frame	Domain	Topology
186	U:433670.3:2000SEP08	75	161	forward 3	TM	N in
186	U:433670.3:2000SEP08	387	449	forward 3	TM	N in
186	U:433670.3:2000SEP08	651	737	forward 3	TM	N in
187	U:202943.4:2000SEP08	535	609	forward 1	TM	N out
187	U:202943.4:2000SEP08	1570	1656	forward 1	TM	N out
187	U:202943.4:2000SEP08	342	395	forward 3	TM	N out
187	U:202943.4:2000SEP08	495	557	forward 3	TM	N out
187	U:202943.4:2000SEP08	579	641	forward 3	TM	N out
187	U:202943.4:2000SEP08	759	845	forward 3	TM	N out
187	U:202943.4:2000SEP08	1653	1721	forward 3	TM	N out
187	U:202943.4:2000SEP08	1905	1991	forward 3	TM	N out
188	U:068682.1:2000SEP08	700	786	forward 1	TM	N out
189	U:203301.3:2000SEP08	643	729	forward 1	TM	N out
189	U:203301.3:2000SEP08	1510	1572	forward 1	TM	N out
189	U:203301.3:2000SEP08	1585	1647	forward 1	TM	N out
189	U:203301.3:2000SEP08	1774	1842	forward 1	TM	N out
189	U:203301.3:2000SEP08	2014	2100	forward 1	TM	N out
189	U:203301.3:2000SEP08	2158	2220	forward 1	TM	N out
189	U:203301.3:2000SEP08	2383	2469	forward 1	TM	N out
189	U:203301.3:2000SEP08	2548	2634	forward 1	TM	N out
189	U:203301.3:2000SEP08	2662	2724	forward 1	TM	N out
189	U:203301.3:2000SEP08	2749	2835	forward 1	TM	N out
189	U:203301.3:2000SEP08	1685	1771	forward 2	TM	N in
189	U:203301.3:2000SEP08	2162	2233	forward 2	TM	N in
189	U:203301.3:2000SEP08	2342	2428	forward 2	TM	N in
189	U:203301.3:2000SEP08	2519	2572	forward 2	TM	N in
189	U:203301.3:2000SEP08	2597	2674	forward 2	TM	N in
189	U:203301.3:2000SEP08	117	170	forward 3	TM	N in
189	U:203301.3:2000SEP08	387	470	forward 3	TM	N in
189	U:203301.3:2000SEP08	1086	1136	forward 3	TM	N in
189	U:203301.3:2000SEP08	1344	1430	forward 3	TM	N in
189	U:203301.3:2000SEP08	1626	1712	forward 3	TM	N in
189	U:203301.3:2000SEP08	2364	2435	forward 3	TM	N in
189	U:203301.3:2000SEP08	2604	2672	forward 3	TM	N in
190	U:020726.3:2000SEP08	1363	1449	forward 1	TM	N in
190	U:020726.3:2000SEP08	1795	1863	forward 1	TM	N in
190	U:020726.3:2000SEP08	1930	1992	forward 1	TM	N in
190	U:020726.3:2000SEP08	545	631	forward 2	TM	N in
190	U:020726.3:2000SEP08	656	742	forward 2	TM	N in
190	U:020726.3:2000SEP08	767	829	forward 2	TM	N in
190	U:020726.3:2000SEP08	839	901	forward 2	TM	N in
190	U:020726.3:2000SEP08	938	1024	forward 2	TM	N in
190	U:020726.3:2000SEP08	1391	1477	forward 2	TM	N in
190	U:020726.3:2000SEP08	1598	1672	forward 2	TM	N in
190	U:020726.3:2000SEP08	1790	1852	forward 2	TM	N in
190	U:020726.3:2000SEP08	1874	1936	forward 2	TM	N in
190	U:020726.3:2000SEP08	1958	2020	forward 2	TM	N in
191	U:027209.1:2000SEP08	25	111	forward 1	TM	N out
191	U:027209.1:2000SEP08	1129	1203	forward 1	TM	N out
191	U:027209.1:2000SEP08	1219	1305	forward 1	TM	N out

147M 200

TABLE 5

SEQ ID NO:	Template ID	Component ID	Start	Stop
189	LI:203301.3:2000SEP08	g4310000	2084	2547
189	LI:203301.3:2000SEP08	2187776T6	2086	2569
189	LI:203301.3:2000SEP08	2187776F6	2093	2584
189	LI:203301.3:2000SEP08	2187776H1	2093	2377
189	LI:203301.3:2000SEP08	956099T6	2098	2568
189	LI:203301.3:2000SEP08	g4683266	2139	2608
189	LI:203301.3:2000SEP08	7247184H1	2140	2722
189	LI:203301.3:2000SEP08	g3240807	2142	2612
189	LI:203301.3:2000SEP08	g3804586	2145	2605
189	LI:203301.3:2000SEP08	g5638660	2154	2602
189	LI:203301.3:2000SEP08	g4297425	2160	2602
189	LI:203301.3:2000SEP08	g4736851	2167	2602
189	LI:203301.3:2000SEP08	2516560T6	2166	2550
189	LI:203301.3:2000SEP08	g7150807	2173	2610
189	LI:203301.3:2000SEP08	g3753686	2240	2617
189	LI:203301.3:2000SEP08	669418H1	2286	2573
189	LI:203301.3:2000SEP08	669124H1	2286	2558
189	LI:203301.3:2000SEP08	g5742262	2297	2602
189	LI:203301.3:2000SEP08	g1220014	2302	2615
189	LI:203301.3:2000SEP08	g4987993	2334	2602
189	LI:203301.3:2000SEP08	g3839358	2344	2602
189	LI:203301.3:2000SEP08	4913428H1	2456	2741
189	LI:203301.3:2000SEP08	3641081H1	2602	2728
189	LI:203301.3:2000SEP08	3485460H1	2614	2890
189	LI:203301.3:2000SEP08	4106769F6	2722	3162
189	LI:203301.3:2000SEP08	4106769H1	2721	2910
190	LI:020726.3:2000SEP08	70858739V1	458	1073
190	LI:020726.3:2000SEP08	71225604V1	499	1009
190	LI:020726.3:2000SEP08	70858001V1	500	1062
190	LI:020726.3:2000SEP08	70796452V1	508	901
190	LI:020726.3:2000SEP08	71226010V1	563	1119
190	LI:020726.3:2000SEP08	7179848H1	562	912
190	LI:020726.3:2000SEP08	71224790V1	581	1124
190	LI:020726.3:2000SEP08	70793655V1	639	743
190	LI:020726.3:2000SEP08	70855458V1	701	1340
190	LI:020726.3:2000SEP08	70858062V1	702	1306
190	LI:020726.3:2000SEP08	71226289V1	729	1174
190	LI:020726.3:2000SEP08	71224784V1	765	1232
190	LI:020726.3:2000SEP08	71224903V1	774	1282
190	LI:020726.3:2000SEP08	70861562V1	837	1436
190	LI:020726.3:2000SEP08	70855548V1	835	1325
190	LI:020726.3:2000SEP08	70855022V1	860	1390
190	LI:020726.3:2000SEP08	7724233H1	854	1272
190	LI:020726.3:2000SEP08	5311769H1	862	1049
190	LI:020726.3:2000SEP08	5311769F8	862	1491
190	LI:020726.3:2000SEP08	70857896V1	913	1567
190	LI:020726.3:2000SEP08	70857337V1	915	1530
190	LI:020726.3:2000SEP08	71225621V1	947	1531
190	LI:020726.3:2000SEP08	71225381V1	981	1564
190	LI:020726.3:2000SEP08	5990468H1	1046	1194

TABLE 5

SEQ ID NO:	Template ID	Component ID	Start	Stop
190	U:020726.3:2000SEP08	6059824F8	1061	1589
190	U:020726.3:2000SEP08	71225618V1	1091	1435
190	U:020726.3:2000SEP08	70858268V1	788	1411
190	U:020726.3:2000SEP08	70858492V1	790	1259
190	U:020726.3:2000SEP08	70796790V1	810	1065
190	U:020726.3:2000SEP08	71227063V1	812	1095
190	U:020726.3:2000SEP08	70858313V1	815	1364
190	U:020726.3:2000SEP08	8097890H1	1	501
190	U:020726.3:2000SEP08	7441541H1	13	484
190	U:020726.3:2000SEP08	70855180V1	777	1400
190	U:020726.3:2000SEP08	70854812V1	781	1360
190	U:020726.3:2000SEP08	71225057V1	314	873
190	U:020726.3:2000SEP08	70856238V1	1167	1506
190	U:020726.3:2000SEP08	7724233J1	1178	1737
190	U:020726.3:2000SEP08	70858568V1	1248	1728
190	U:020726.3:2000SEP08	70856388V1	1272	1750
190	U:020726.3:2000SEP08	2738605T6	1334	1877
190	U:020726.3:2000SEP08	71224942V1	1370	1816
190	U:020726.3:2000SEP08	71224964V1	1376	1874
190	U:020726.3:2000SEP08	2738601T7	1383	1810
190	U:020726.3:2000SEP08	5906145H1	1506	1811
190	U:020726.3:2000SEP08	5906145F6	1506	1790
190	U:020726.3:2000SEP08	g1383637	1553	1914
190	U:020726.3:2000SEP08	7427246H1	1717	2139
191	U:027209.1:2000SEP08	1388139H1	1	262
191	U:027209.1:2000SEP08	2737908H1	3	238
191	U:027209.1:2000SEP08	2737908F6	3	395
191	U:027209.1:2000SEP08	6306041F8	5	651
191	U:027209.1:2000SEP08	6306041H1	5	464
191	U:027209.1:2000SEP08	70563274V1	229	974
191	U:027209.1:2000SEP08	70563246V1	340	843
191	U:027209.1:2000SEP08	70565819V1	440	1018
191	U:027209.1:2000SEP08	1904413H1	563	709
191	U:027209.1:2000SEP08	70563142V1	627	1120
191	U:027209.1:2000SEP08	70564681V1	633	1098
191	U:027209.1:2000SEP08	70565477V1	675	1279
191	U:027209.1:2000SEP08	70562989V1	740	1342
191	U:027209.1:2000SEP08	70563804V1	745	1341
191	U:027209.1:2000SEP08	70565738V1	758	1337
191	U:027209.1:2000SEP08	2739431H1	801	1034
191	U:027209.1:2000SEP08	2739431F6	801	1135
191	U:027209.1:2000SEP08	70564623V1	974	1362
191	U:027209.1:2000SEP08	6306041T8	982	1512
191	U:027209.1:2000SEP08	2737908T6	1116	1572
191	U:027209.1:2000SEP08	1265467H1	1494	1829
191	U:027209.1:2000SEP08	2819460T6	1496	1585
191	U:027209.1:2000SEP08	2819460F6	1503	1615
191	U:027209.1:2000SEP08	2819460H1	1504	1587
192	U:108819.1:2000SEP08	71304273V1	726	1132
192	U:108819.1:2000SEP08	2992317H1	831	1113

TABLE 6

SEQ ID NO:	Template ID	Tissue Distribution
180	LI:451710.1:2000SEP08	Connective Tissue - 90%, Nervous System - 10%
181	LI:406882.2:2000SEP08	Sense Organs - 34%, Unclassified/Mixed - 15%, Nervous System - 15%
182	LI:728223.1:2000SEP08	Nervous System - 100%
183	LI:289783.19:2000SEP08	Digestive System - 38%, Unclassified/Mixed - 15%, Respiratory System - 12%
184	LI:235255.8:2000SEP08	Exocrine Glands - 23%, Connective Tissue - 12%, Female Genitalia - 11%
185	LI:237693.5:2000SEP08	Urinary Tract - 32%, Hemic and Immune System - 25%, Endocrine System - 21%
186	LI:433670.3:2000SEP08	Nervous System - 50%, Male Genitalia - 25%, Digestive System - 25%
187	LI:202943.4:2000SEP08	Embryonic Structures - 42%, Liver - 19%, Unclassified/Mixed - 16%
188	LI:068682.1:2000SEP08	Unclassified/Mixed - 47%, Digestive System - 21%, Male Genitalia - 19%
189	LI:203301.3:2000SEP08	Germ Cells - 19%, Male Genitalia - 14%, Respiratory System - 12%
190	LI:020726.3:2000SEP08	Sense Organs - 79%
191	LI:027209.1:2000SEP08	Musculoskeletal System - 53%, Female Genitalia - 27%, Exocrine Glands - 17%
192	LI:108819.1:2000SEP08	Urinary Tract - 45%, Digestive System - 28%, Skin - 15%
193	LI:021759.1:2000SEP08	Hemic and Immune System - 40%, Nervous System - 22%, Liver - 11%
194	LI:1165967.1:2000SEP08	Liver - 50%, Stomatognathic System - 50%
195	LI:1166315.1:2000SEP08	Cardiovascular System - 33%, Urinary Tract - 27%, Female Genitalia - 20%, Hemic and Immune System - 20%
196	LI:204626.1:2000SEP08	Digestive System - 58%, Nervous System - 15%, Exocrine Glands - 13%
197	LI:801140.1:2000SEP08	Embryonic Structures - 50%, Cardiovascular System - 38%
198	LI:286639.1:2000SEP08	Germ Cells - 64%
199	LI:288905.4:2000SEP08	Unclassified/Mixed - 76%, Nervous System - 24%
200	LI:332161.1:2000SEP08	Cardiovascular System - 47%, Nervous System - 14%, Connective Tissue - 10%
201	LI:184867.1:2000SEP08	Unclassified/Mixed - 33%, Embryonic Structures - 24%, Male Genitalia - 16%
202	LI:229932.4:2000SEP08	Musculoskeletal System - 40%, Cardiovascular System - 21%
203	LI:1189932.1:2000SEP08	Embryonic Structures - 37%, Germ Cells - 18%, Sense Organs - 13%
204	LI:1076689.1:2000SEP08	Liver - 99%
205	LI:415181.2:2000SEP08	Nervous System - 100%



TABLE 7

SEQ ID NO:	Frame	Length	Start	Stop	GI Number	Probability Score	Annotation
442	2	311	743	1675	g13278193	1.00E-153	Similar to EHM2 gene
442	2	311	743	1675	g10434740	1.00E-140	unnamed protein product
443	2	529	416	2002	g12834087	1.00E-155	putative
443	2	529	416	2002	g2463628	6.00E-46	putative monocarboxylate transporter
443	2	529	416	2002	g7328162	1.00E-40	hypothetical protein
444	1	335	4	1008	g13159480	1.00E-128	Translation may initiate at the ATG codon at nucleotides 40-42 or the ATG at nucleotides 43-45
444	1	335	4	1008	g9229906	5.00E-35	fibrinogen-like protein
444	1	335	4	1008	g387156	8.00E-33	fibrinogen-like protein
445	1	329	196	1182	g4582324	0	dJ708F5.1 (PUTATIVE novel Collagen alpha 1 LIKE protein)
445	1	329	196	1182	g12052774	0	hypothetical protein
445	1	329	196	1182	g2326442	1.00E-38	collagen type XI alpha 1 chain
446	2	395	110	1294	g12642596	0	nuclear receptor co-repressor/HDAC3 complex subunit TBLR1
446	2	395	110	1294	g10434648	0	unnamed protein product
446	2	395	110	1294	g12006104	0	IRA1
447	1	163	1	489	g12858551	7.00E-67	putative
447	1	163	1	489	g12805285	7.00E-67	Similar to ribosomal protein S27a
447	1	163	1	489	g1050756	7.00E-67	fusion protein: ubiquitin (bases 43_513); ribosomal protein S27a (bases 217_532)
448	3	78	330	563	g8699209	5.00E-06	cyclophilin A
448	3	78	330	563	g50621	5.00E-06	cyclophilin (AA 1 - 164)
448	3	78	330	563	g49496	5.00E-06	cyclophilin (AA 1-164)
449	1	314	277	1218	g12844770	1.00E-130	putative
449	1	314	277	1218	g12861366	1.00E-127	putative
449	1	314	277	1218	g12857383	7.00E-50	putative
450	1	130	181	570	g2843171	4.00E-06	zinc finger protein
450	1	130	181	570	g5817149	5.00E-06	hypothetical protein
450	1	130	181	570	g10434142	5.00E-06	unnamed protein product
451	1	176	697	1224	g13383265	1.00E-63	actin related protein
451	1	176	697	1224	g13938319	3.00E-62	Unknown (protein for MGC:15664)
451	1	176	697	1224	g12840619	2.00E-51	putative
452	3	671	3	2015	g12698057	1.00E-168	KIAA1756 protein

## CLAIMS

## What is claimed is:

- 5           1. An isolated polynucleotide selected from the group consisting of:
- a)     a polynucleotide comprising a polynucleotide sequence selected from the group  
              consisting of SEQ ID NO:1-252,
- b)     a polynucleotide comprising a naturally occurring polynucleotide sequence at least  
              90% identical to a polynucleotide sequence selected from the group consisting of SEQ  
10           ID NO:1-252,
- c)     a polynucleotide complementary to the polynucleotide of a),
- d)     a polynucleotide complementary to the polynucleotide of b), and
- e)     an RNA equivalent of a)-d).
- 15           2. An isolated polynucleotide of claim 1, selected from the group consisting of SEQ ID NO:1-  
252.
3. An isolated polynucleotide comprising at least 30 contiguous nucleotides of a polynucleotide  
of claim 1.
- 20           4. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide  
of claim 1.
5. A composition for the detection of expression of disease detection and treatment  
25           polynucleotides comprising at least one of the polynucleotides of claim 1 and a detectable label.
6. A method for detecting a target polynucleotide in a sample, said target polynucleotide  
having a sequence of a polynucleotide of claim 1, the method comprising:
- a)     amplifying said target polynucleotide or fragment thereof using polymerase chain  
30           reaction amplification, and
- b)     detecting the presence or absence of said amplified target polynucleotide or fragment  
thereof, and, optionally, if present, the amount thereof.
7. A method for detecting a target polynucleotide in a sample, said target polynucleotide  
35           having a sequence of a polynucleotide of claim 1, the method comprising:

- 5
- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
  - b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

10 8. A method of claim 7, wherein the probe comprises at least 30 contiguous nucleotides.

9. A method of claim 7, wherein the probe comprises at least 60 contiguous nucleotides.

15 10. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 1.

11. A cell transformed with a recombinant polynucleotide of claim 10.

12. A transgenic organism comprising a recombinant polynucleotide of claim 10.

20 13. A method for producing a disease detection and treatment polypeptide encoded by a polynucleotide of claim 1, the method comprising:

- a) culturing a cell under conditions suitable for expression of the disease detection and treatment polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 1, and
  - b) recovering the disease detection and treatment polypeptide so expressed.
- 25

30 14. A method of claim 13, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:253-506.

15. An isolated disease detection and treatment polypeptide (MDDT) encoded by at least one of the polynucleotides of claim 2.

35 16. A method of screening for a test compound that specifically binds to the polypeptide of claim 15, the method comprising:

- a) combining the polypeptide of claim 15 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 15 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 15.

5

17. A microarray wherein at least one element of the microarray is a polynucleotide of claim 3.

18. A method for generating a transcript image of a sample which contains polynucleotides, the method comprising:

10

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 17 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

15

19. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence of a polynucleotide of claim 1, the method comprising:

20

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

25

20. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 1 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 1 or fragment thereof,
- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a

30

35

difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

21. An array comprising different nucleotide molecules affixed in distinct physical locations  
5 on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 1.

22. An array of claim 21, wherein said first oligonucleotide or polynucleotide sequence is  
10 completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

23. An array of claim 21, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide

24. An array of claim 21, wherein said first oligonucleotide or polynucleotide sequence is  
15 completely complementary to said target polynucleotide.

25. An array of claim 21, which is a microarray.

26. An array of claim 21, further comprising said target polynucleotide hybridized to a  
20 nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

27. An array of claim 21, wherein a linker joins at least one of said nucleotide molecules to  
said solid substrate.

28. An array of claim 21, wherein each distinct physical location on the substrate contains  
multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical  
location have the same sequence, and each distinct physical location on the substrate contains  
nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at  
30 another distinct physical location on the substrate.

29. An isolated polypeptide selected from the group consisting of:

- a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:253-506,
- 35 b) a polypeptide comprising a naturally occurring amino acid sequence at least 90%

identical to an amino acid sequence selected from the group consisting of SEQ ID NO:253-506,

- c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, and
- d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506.

30. An isolated polypeptide of claim 29, having a sequence selected from the group consisting of SEQ ID NO:253-506.

31. An isolated polynucleotide encoding a polypeptide of claim 29.

32. An isolated polynucleotide encoding a polypeptide of claim 30.

33. An isolated polynucleotide of claim 32, having a sequence selected from the group consisting of SEQ ID NO:1-252.

34. An isolated antibody which specifically binds to a disease detection and treatment polypeptide of claim 29.

35. A diagnostic test for a condition or disease associated with the expression of MDDT in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 34, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

36. The antibody of claim 34, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

37. A composition comprising an antibody of claim 34 and an acceptable excipient.

38. A method of diagnosing a condition or disease associated with the expression of MDDT in a subject, comprising administering to said subject an effective amount of the composition of claim 37.

39. A composition of claim 37, wherein the antibody is labeled.

40. A method of diagnosing a condition or disease associated with the expression of MDDT in a subject, comprising administering to said subject an effective amount of the composition of claim 39.

41. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 34, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506.

42. An antibody produced by a method of claim 41.

43. A composition comprising the antibody of claim 42 and a suitable carrier.

44. A method of making a monoclonal antibody with the specificity of the antibody of claim 34, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506.

5

45. A monoclonal antibody produced by a method of claim 44.

46. A composition comprising the antibody of claim 45 and a suitable carrier.

10

47. The antibody of claim 34, wherein the antibody is produced by screening a Fab expression library.

48. The antibody of claim 34, wherein the antibody is produced by screening a recombinant immunoglobulin library.

15

49. A method of detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506 in a sample, the method comprising:

- a) incubating the antibody of claim 34 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506 in the sample.

20

50. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506 from a sample, the method comprising:

25

- a) incubating the antibody of claim 34 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:253-506.

30

51. A composition comprising a polypeptide of claim 29 and a pharmaceutically acceptable excipient.

52. A composition of claim 51, wherein the polypeptide has an amino acid sequence of SEQ ID NO:253-506.

35



53. A method for treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment the composition of claim 51.

5 54. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 29, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 29 to a compound, and
- b) detecting agonist activity in the sample.

10 55. A composition comprising an agonist compound identified by a method of claim 54 and a pharmaceutically acceptable excipient.

15 56. A method for treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment a composition of claim 55.

57. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 29, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 29 to a compound, and
- 20 b) detecting antagonist activity in the sample.

58. A composition comprising an antagonist compound identified by a method of claim 57 and a pharmaceutically acceptable excipient.

25 59. A method for treating a disease or condition associated with overexpression of functional MDDT, comprising administering to a patient in need of such treatment a composition of claim 58.

60. A method of screening for a compound that modulates the activity of the polypeptide of claim 29, said method comprising:

- 30 a) combining the polypeptide of claim 29 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 29;
- b) assessing the activity of the polypeptide of claim 29 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 29 in the presence of the test  
35 compound with the activity of the polypeptide of claim 29 in the absence of the test

```

ccttttagtat gttcacacat caccagact gtatctcagg gagaagggtt gtgtttgtga 2040
acaagggtgcc cattattccc ccaccacatg ccatccaaag agatcgaggt taaaagaatc 2100
agacacacta aaaacacagt ggaagcatgt ctaggcataa tatatcagcc ttttgagata 2160
tttattaaaa acttgttctt atacactatt ttaatatgtc taaagtgaca cataccctta 2220
atttaagtgc ctaagatttt aattaatata atttttaatt acaaacactt ttacttttta 2280
ataaacacga atattttaac ttaccaaag atttagagat cacattaaat aatgtggctc 2340
atgtatagag aaatctgaat gtttaggttt tgaatggtta actgattaac attcagttat 2400
cataacagtt aaccagtggt taactgaatg gttacattac tactgaatgg ttcatataaa 2460
atgaaaaaca taatattaca ttgtctctt catttgactg gcttttcac atattcatca 2520
tcaaacacct gacagattac ccaggtttgc aacagtctca taatctgac ctgtataacc 2580
tttaaaataa aattctgaag ccaagaacca tatgattttt agattgtatt actatttttg 2640
atccttttta aacttttagaa tttcaagctc cttggtttac ttaatagaag tgaattagct 2700
ggaaaaagtt catttctgat aggaaaacat gtttaaggga gcatttcttc tgtctttttt 2760
tagttttctt ttatatttta gaaaaatttt caaattttta aactatgtac atagaaattt 2820
atagatattt tatatttcga agagtttata atgcttcaag ataatgaagc ttcatataaa 2880
acagcttaac tcatttaatg catcctgctc acatagaggt tggttttata tctcaaatcc 2940
ttcctgccag acaaagatta aggtttcaga gggcttaata ctggactgtt ggtttaatac 3000
tgaatggttg gttgcagttt gtccatagaa gacctaata agcagaactg actaatattt 3060
aaagaatgaa atttgtaggc aggttcaatc tgagtttagt ataattatgt gtaagaggat 3120
ggccactcaa agtatttagc acatcctcta aatataaata ga 3162

```

&lt;210&gt; 190

&lt;211&gt; 2139

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: LI:020726.3:2000SEP08

&lt;400&gt; 190

```

gggcagggtg taggtttggg cacctccaaa cgcaccggtt ttaaagccac ctggactgag 60
gcgtcgagct ttcagctcca ccaaacgctc acctggcctg gcagcgagcg gcggaagagc 120
ccgggagccc ctcacagagc gcaccgagcc gggcggagag ctgagccgca ggcacccgcg 180
tctccaggat gataggcgac attgcaacaa atctctacac ccagcagctc aggggggctcc 240
aagcagagca gcaagttcga ggatccggcg gtggagccga gtgagggcgc agcccaacgg 300
gcctcggggc ctcagctcag ctccgctccc cagccttctc cgccggcagcc tcttcagcct 360
gctggccgca agtgccgccc ctaaaggccc caaatgccct gtacacacca ggtgaagagc 420
gcggaagcgc ctgcagagca gaattaaaga aaaatccttg aaaatgtata ccagtcatga 480
agatattggg tatgattttg aagatggccc caaagacaaa aagacactga agccccaccc 540
aaacattgat gcgggatggg cttggatgat ggtgctctcc tctttctttg tgcacatcct 600
catcatgggc tccagatagg ccctgggtgt cctcaacgtg gaatggctgg aagaattcca 660
ccagagccgc ggcctgaccg cctgggtcag ctccctcagc atgggcatca ccttgatagt 720
gggccccttc atcggcctgt tcattaacac ctgtgggtgc cgcagactg cgatcattgg 780
agggtcgtc aactccctgg gctgggtgtg gagtgcctat gctgcaaacg tgcattatct 840
cttcattact tttggagtcg cagctggcct gggcagcggg atggcctacc tgccagcgtt 900
ggtcatgggt ggcaggattt tccagaagag acgcgcctc gccagggcc tcagcaccac 960
ggggaccgga ttcggtacgt tccaatgac tgtgctgtg aagtacctgt gcgcagagta 1020
cggctggagg aatgccatgt tgatccaagg tgccgtttcc ctaaacctgt gtgtttgtgg 1080
ggcgtcctat agggccctct ctccctgtaa aaacccaaac gaccaggag agaaagatgt 1140
gcgtggcctg ccagcgact ccacagaatc tgtgaagtca actggacagc agggaagaac 1200
agaagagaag gatggtgggc tcgggaacga ggagaccctc tgcgacctgc aagcccagga 1260
gtgccccgat caggccgggc acaggaagaa catgtgtgcc ctccggattc tgaagactgt 1320
cagctggctc acctagagag tcaggaaggg cttcgaggac tggatttcgg gctactttgg 1380
gacagcctct ctatttacia atcgaatggt ttagccttt attttctggg ctttgtttgc 1440
atacagcagc ttgtctatcc ccttcattca cctcccagaa atcgtcaatt tgtataactt 1500
atcggagcaa aacgacgttt tccctctgac gtcaattata gcaatagttc acatctttgg 1560
aaaagtgate ctgggctgca tagccgactt gccttgcaat agtgtttgga atgtcttcct 1620
gttgccaac ttcaccctg tctcagtat ttttattctg ccgttgatgc acacgtacgc 1680
tggctggcg gtcattctgt cgctgatagg gttttccagt ggttatttct ccctaagcc 1740
cgtagtgact gaagacttgg ttggcattga acacctggcc aatgcctacg gcatcatcat 1800
ctgtgctaata ggcattctct cattgtctgg accaccttt gcagggtgga tctatgacat 1860
cacgcaaaaa tatgattttt ccttctacat atgtggtttg ctttacctga taggaatact 1920

```

```

ctttttactt attcagccgt gcattcgaat tatagaacat ccagaagaaa atacatggat 1980
gggtcacatg ttttagtatca tgtaatgttc cgtgtagggt tcattgtaat actcatgcct 2040
acctcgcatg gttgctgtga ggcacctatg acaggacgtg ggaaagcatt ttggtacggg 2100
aactggcact gtcatttgta aatgccattt gtcacagcc 2139

```

&lt;210&gt; 191

&lt;211&gt; 1615

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: LI:027209.1:2000SEP08

&lt;400&gt; 191

```

tgatatttga agaagtgttt tcatctatcc aagaaaaata tgatgtctcc atcccaagcc 60
tcactcttat tcttaaatgt atgtattttt atttgtggag aagctgtaca aggtaactgt 120
gtacatcatt ctacggactc ttcagtagtt aacattgtag aagatggatc taatgcaaaa 180
gatgaaagta aaagtaatga tactgtttgt aagggaagact gtgaggaatc atgtgatgtt 240
aaaactaaaa ttacacgaga agaaaaacat ttcatgtgta gaaatttgca aaattctatt 300
gtttcctaca caagaagtac caaaaaacta ctaagggaata tgatggatga gcaacaagct 360
tccttggatt atttatctaa tcaggtaatg tgtgacatgg attacagagg aggtggatgg 420
actgtgatac agaaaaagaat tgatgggata attgatctcc agaggttgtg gtgtgattat 480
ctggatggat ttggagatct tctaggagaa ttttggctag gactgaaaaa gattttttat 540
atagtaaadc agaaaaatc cagttttatg ctgtatgtgg ctttggatc tgaagatgac 600
actcttgctt atgcatcata tgataatttt tggctagagg atgaaacgag atttttttaa 660
atgcacttag gacggtattc aggaatgct ggtgatgcat tccggggtct caaaaaagaa 720
gataatcaaa atgcaatgcc ttttagcaca tcagatgttg ataatgatgg gtgtcgccct 780
gcatgcctgg tcaatggtca gtctgtgaag agctgcagtc acctccataa caagaccggc 840
tggtgggtta acgagtgtgg tctagcaaat ctaaattggca ttcattcact ctctggaaaa 900
ttgcttgcaa ctggaattca atggggcacg tggaccaaaa acaactcacc tgtcaagatt 960
aaatctgttt caatgaaaat tagaagaatg tacaatccat attttaata atctcattta 1020
acattgtaat gcaagttcta caatgataat atattaaaga tttttaaaag tttatctttt 1080
cacttagtgt ttcaaacata ttaggcacaaa tttactgta gatggcattt agatgttatg 1140
agtttaatta gaaaacttca attttgtagt attctataaa agaaaacatg gcttattgta 1200
tgtttttact tctgactata ttaacaatat acaatgaaat ttgtttcaag tgaactacaa 1260
cttgtcttcc taaaatttat agtgatttta aaggattttg ccttttcttt gaagcatttt 1320
taaaccataa tatgttgtaa ggaaaattga agggaatatt ttacttattt ttatacttta 1380
tatgattata taatctacag ataatttcta ctgaagacag ttacaataaa taactttatg 1440
cagattaata tataagctac acatgatgta aaaaccttac tatttctagg tgatgccata 1500
ccattttaaa agtagtaaga gtttgctgcc caaatagttt ttcttggttt catatcta 1560
catggttaac tattttgtta ttgtttgtaa taaatatatg tacttttata tcctg 1615

```

&lt;210&gt; 192

&lt;211&gt; 1184

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: LI:108819.1:2000SEP08

&lt;400&gt; 192

```

ggaattcggc tcgagcgtcg caggagagaac ggactccggg cggagggcag ccaatccgtt 60
tcagcgcagg tcttgctcgg gttgggcttg ccactgcctg gaacatacct gtccccctgg 120
cgcaacactc agctggctgc gacggcaacc ccgagcctgg aactgcccgc aggaatccta 180
aaacccaaaat attagaacga aaacagaaac atgggtcact atattacatt tctctgcatt 240
gttttgggtg tgcttcttca gaattctgtg ttagctgaag atggggaagt aagatcaagt 300
tgtcgtagtg ctccgacaga tttagtcttc atcttagatg gctcttatag tgttgccca 360
gaaaactttg aaatagtga aaagtggctt gtcaatatca caaaaaactt tgacataggg 420
ccgaagttaa ttcaagtgg agtggttcaa tatagtgaat accctgtgct ggagattcct 480
ctcggaagct atgattcagg agaacatttg acggcagcag tggaaatccat actctactta 540
ggaggaaaca caaagacagg gaaggccatc cagtttgcgc tcgattacct ttttgccaag 600

```

<210> 442  
 <211> 311  
 <212> PRT  
 <213> Homo sapiens  
  
 <220>  
 <221> misc\_feature  
 <223> Incyte ID No: LI:203301.3.orf2:2000SEP08

<400> 442  
 Thr Cys Pro Arg Arg His Tyr Ser Cys Leu Val Gln Thr Glu Glu  
 1 5 10 15  
 Met Glu Leu Ala Ile Phe Glu Lys Trp Lys Glu Tyr Arg Gly Gln  
 20 25 30  
 Thr Pro Ala Gln Ala Glu Thr Asn Tyr Leu Asn Lys Ala Lys Trp  
 35 40 45  
 Leu Glu Met Tyr Gly Val Asp Met His Val Lys Gly Arg Asp  
 50 55 60  
 Gly Asn Asp Tyr Ser Leu Gly Leu Thr Pro Thr Gly Val Leu Val  
 65 70 75  
 Phe Glu Gly Asp Thr Lys Ile Gly Leu Phe Phe Trp Pro Lys Ile  
 80 85 90  
 Thr Arg Leu Asp Phe Lys Lys Asn Lys Leu Thr Leu Val Val Val  
 95 100 105  
 Glu Asp Asp Asp Gln Gly Lys Glu Gln His Thr Phe Val Phe  
 110 115 120  
 Arg Leu Asp His Pro Lys Ala Cys Lys His Leu Trp Lys Cys Ala  
 125 130 135  
 Val Glu His His Ala Phe Phe Arg Leu Arg Gly Pro Val Gln Lys  
 140 145 150  
 Ser Ser His Arg Ser Gly Phe Ile Arg Leu Gly Ser Arg Phe Arg  
 155 160 165  
 Tyr Ser Gly Lys Thr Glu Tyr Gln Thr Thr Lys Thr Asn Lys Ala  
 170 175 180  
 Arg Arg Ser Thr Ser Phe Glu Arg Arg Pro Ser Lys Arg Tyr Ser  
 185 190 195  
 Arg Arg Thr Leu Gln Met Lys Ala Cys Ala Thr Lys Pro Glu Glu  
 200 205 210  
 Leu Ser Val His Asn Asn Val Ser Thr Gln Ser Asn Gly Ser Gln  
 215 220 225  
 Gln Ala Trp Gly Met Arg Ser Ala Leu Pro Val Ser Pro Ser Ile  
 230 235 240  
 Ser Ser Ala Pro Val Pro Val Glu Ile Glu Asn Leu Pro Gln Ser  
 245 250 255  
 Pro Gly Thr Asp Gln His Asp Arg Lys Trp Leu Ser Ala Ala Ser  
 260 265 270  
 Asp Cys Cys Gln Arg Gly Gly Asn Gln Trp Asn Thr Arg Ala Leu  
 275 280 285  
 Pro Pro Pro Gln Thr Ala His Arg Asn Tyr Thr Asp Phe Val His  
 290 295 300  
 Glu His Asn Val Lys Asn Ala Gly Ile Pro Ser  
 305 310

<210> 443  
 <211> 529  
 <212> PRT  
 <213> Homo sapiens  
  
 <220>  
 <221> misc\_feature  
 <223> Incyte ID No: LI:020726.3.orf2:2000SEP08

&lt;400&gt; 443

Arg	Ala	Arg	Lys	Arg	Leu	Gln	Ser	Arg	Ile	Lys	Glu	Lys	Ser	Trp
1				5					10					15
Lys	Met	Tyr	Thr	Ser	His	Glu	Asp	Ile	Gly	Tyr	Asp	Phe	Glu	Asp
				20					25					30
Gly	Pro	Lys	Asp	Lys	Lys	Thr	Leu	Lys	Pro	His	Pro	Asn	Ile	Asp
				35					40					45
Gly	Gly	Trp	Ala	Trp	Met	Met	Val	Leu	Ser	Ser	Phe	Phe	Val	His
				50					55					60
Ile	Leu	Ile	Met	Gly	Ser	Gln	Met	Ala	Leu	Gly	Val	Leu	Asn	Val
				65					70					75
Glu	Trp	Leu	Glu	Glu	Phe	His	Gln	Ser	Arg	Gly	Leu	Thr	Ala	Trp
				80					85					90
Val	Ser	Ser	Leu	Ser	Met	Gly	Ile	Thr	Leu	Ile	Val	Gly	Pro	Phe
				95					100					105
Ile	Gly	Leu	Phe	Ile	Asn	Thr	Cys	Gly	Cys	Arg	Gln	Thr	Ala	Ile
				110					115					120
Ile	Gly	Gly	Leu	Val	Asn	Ser	Leu	Gly	Trp	Val	Leu	Ser	Ala	Tyr
				125					130					135
Ala	Ala	Asn	Val	His	Tyr	Leu	Phe	Ile	Thr	Phe	Gly	Val	Ala	Ala
				140					145					150
Gly	Leu	Gly	Ser	Gly	Met	Ala	Tyr	Leu	Pro	Ala	Val	Val	Met	Val
				155					160					165
Gly	Arg	Tyr	Phe	Gln	Lys	Arg	Arg	Ala	Leu	Ala	Gln	Gly	Leu	Ser
				170					175					180
Thr	Thr	Gly	Thr	Gly	Phe	Gly	Thr	Phe	Leu	Met	Thr	Val	Leu	Leu
				185					190					195
Lys	Tyr	Leu	Cys	Ala	Glu	Tyr	Gly	Trp	Arg	Asn	Ala	Met	Leu	Ile
				200					205					210
Gln	Gly	Ala	Val	Ser	Leu	Asn	Leu	Cys	Val	Cys	Gly	Ala	Leu	Met
				215					220					225
Arg	Pro	Leu	Ser	Pro	Gly	Lys	Asn	Pro	Asn	Asp	Pro	Gly	Glu	Lys
				230					235					240
Asp	Val	Arg	Gly	Leu	Pro	Ala	His	Ser	Thr	Glu	Ser	Val	Lys	Ser
				245					250					255
Thr	Gly	Gln	Gln	Gly	Arg	Thr	Glu	Gly	Lys	Asp	Gly	Gly	Leu	Gly
				260					265					270
Asn	Glu	Glu	Thr	Leu	Cys	Asp	Leu	Gln	Ala	Gln	Glu	Cys	Pro	Asp
				275					280					285
Gln	Ala	Gly	His	Arg	Lys	Asn	Met	Cys	Ala	Leu	Arg	Ile	Leu	Lys
				290					295					300
Thr	Val	Ser	Trp	Leu	Thr	Met	Arg	Val	Arg	Lys	Gly	Phe	Glu	Asp
				305					310					315
Trp	Tyr	Ser	Gly	Tyr	Phe	Gly	Thr	Ala	Ser	Leu	Phe	Thr	Asn	Arg
				320					325					330
Met	Phe	Val	Ala	Phe	Ile	Phe	Trp	Ala	Leu	Phe	Ala	Tyr	Ser	Ser
				335					340					345
Phe	Val	Ile	Pro	Phe	Ile	His	Leu	Pro	Glu	Ile	Val	Asn	Leu	Tyr
				350					355					360
Asn	Leu	Ser	Glu	Gln	Asn	Asp	Val	Phe	Pro	Leu	Thr	Ser	Ile	Ile
				365					370					375
Ala	Ile	Val	His	Ile	Phe	Gly	Lys	Val	Ile	Leu	Gly	Val	Ile	Ala
				380					385					390
Asp	Leu	Pro	Cys	Ile	Ser	Val	Trp	Asn	Val	Phe	Leu	Leu	Ala	Asn
				395					400					405
Phe	Thr	Leu	Val	Leu	Ser	Ile	Phe	Ile	Leu	Pro	Leu	Met	His	Thr
				410					415					420
Tyr	Ala	Gly	Leu	Ala	Val	Ile	Cys	Ala	Leu	Ile	Gly	Phe	Ser	Ser
				425					430					435
Gly	Tyr	Phe	Ser	Leu	Met	Pro	Val	Val	Thr	Glu	Asp	Leu	Val	Gly
				440					445					450
Ile	Glu	His	Leu	Ala	Asn	Ala	Tyr	Gly	Ile	Ile	Ile	Cys	Ala	Asn
				455					460					465

Gly	Ile	Ser	Ala	Leu	Leu	Gly	Pro	Pro	Phe	Ala	Gly	Trp	Ile	Tyr	
				470					475					480	
Asp	Ile	Thr	Gln	Lys	Tyr	Asp	Phe	Ser	Phe	Tyr	Ile	Cys	Gly	Leu	
				485					490					495	
Leu	Tyr	Met	Ile	Gly	Ile	Leu	Phe	Leu	Leu	Ile	Gln	Pro	Cys	Ile	
				500					505					510	
Arg	Ile	Ile	Glu	His	Pro	Glu	Glu	Asn	Thr	Trp	Met	Val	His	Met	
				515					520					525	
Phe	Ser	Ile	Met												

&lt;210&gt; 444

&lt;211&gt; 335

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: LI:027209.1.orf1:2000SEP08

&lt;400&gt; 444

Tyr	Leu	Lys	Lys	Cys	Phe	His	Leu	Ser	Lys	Lys	Asn	Met	Met	Ser	
1				5					10					15	
Pro	Ser	Gln	Ala	Ser	Leu	Leu	Phe	Leu	Asn	Val	Cys	Ile	Phe	Ile	
				20					25					30	
Cys	Gly	Glu	Ala	Val	Gln	Gly	Asn	Cys	Val	His	His	Ser	Thr	Asp	
				35					40					45	
Ser	Ser	Val	Val	Asn	Ile	Val	Glu	Asp	Gly	Ser	Asn	Ala	Lys	Asp	
				50					55					60	
Glu	Ser	Lys	Ser	Asn	Asp	Thr	Val	Cys	Lys	Glu	Asp	Cys	Glu	Glu	
				65					70					75	
Ser	Cys	Asp	Val	Lys	Thr	Lys	Ile	Thr	Arg	Glu	Glu	Lys	His	Phe	
				80					85					90	
Met	Cys	Arg	Asn	Leu	Gln	Asn	Ser	Ile	Val	Ser	Tyr	Thr	Arg	Ser	
				95					100					105	
Thr	Lys	Lys	Leu	Leu	Arg	Asn	Met	Met	Asp	Glu	Gln	Gln	Ala	Ser	
				110					115					120	
Leu	Asp	Tyr	Leu	Ser	Asn	Gln	Val	Met	Cys	Asp	Met	Asp	Tyr	Arg	
				125					130					135	
Gly	Gly	Gly	Trp	Thr	Val	Ile	Gln	Lys	Arg	Ile	Asp	Gly	Ile	Ile	
				140					145					150	
Asp	Phe	Gln	Arg	Leu	Trp	Cys	Asp	Tyr	Leu	Asp	Gly	Phe	Gly	Asp	
				155					160					165	
Leu	Leu	Gly	Glu	Phe	Trp	Leu	Gly	Leu	Lys	Lys	Ile	Phe	Tyr	Ile	
				170					175					180	
Val	Asn	Gln	Lys	Asn	Thr	Ser	Phe	Met	Leu	Tyr	Val	Ala	Leu	Glu	
				185					190					195	
Ser	Glu	Asp	Asp	Thr	Leu	Ala	Tyr	Ala	Ser	Tyr	Asp	Asn	Phe	Trp	
				200					205					210	
Leu	Glu	Asp	Glu	Thr	Arg	Phe	Phe	Lys	Met	His	Leu	Gly	Arg	Tyr	
				215					220					225	
Ser	Gly	Asn	Ala	Gly	Asp	Ala	Phe	Arg	Gly	Leu	Lys	Lys	Glu	Asp	
				230					235					240	
Asn	Gln	Asn	Ala	Met	Pro	Phe	Ser	Thr	Ser	Asp	Val	Asp	Asn	Asp	
				245					250					255	
Gly	Cys	Arg	Pro	Ala	Cys	Leu	Val	Asn	Gly	Gln	Ser	Val	Lys	Ser	
				260					265					270	
Cys	Ser	His	Leu	His	Asn	Lys	Thr	Gly	Trp	Trp	Phe	Asn	Glu	Cys	
				275					280					285	
Gly	Leu	Ala	Asn	Leu	Asn	Gly	Ile	His	His	Phe	Ser	Gly	Lys	Leu	
				290					295					300	
Leu	Ala	Thr	Gly	Ile	Gln	Trp	Gly	Thr	Trp	Thr	Lys	Asn	Asn	Ser	
				305					310					315	

## INTERNATIONAL SEARCH REPORT

 International Application No  
 PCT/US 01/27628

## A. CLASSIFICATION OF SUBJECT MATTER

 IPC 7 C12N15/12 C07K14/47 C12Q1/68 A61K31/7088 A01K67/027  
 C12N5/10 C07K16/18 G01N33/68 A61K39/395 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K C12Q G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EMBL, EPO-Internal, SEQUENCE SEARCH, WPI Data, BIOSIS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EMMANUEL DIAS NETO ET AL: "SHOTGUN SEQUENCING OF THE HUMAN TRANSCRIPTOME WITH ORF EXPRESSED SEQUENCE TAGS" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 97, no. 7, 28 March 2000 (2000-03-28), pages 3491-3496, XP000996193 ISSN: 0027-8424 the whole document - & DATABASE EMBL [Online] 15 January 2001 (2001-01-15) DIAS NETO E. ET AL: "RC1-IT025-201100-021-d12 IT0025 Homo sapiens cDNA, mRNA sequence" retrieved from EBI, HINXTON, UK Database accession no. BF770200 XP002219868 -/--	1

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

14 November 2002

Date of mailing of the international search report

17. 02 2003

Name and mailing address of the ISA

 European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

LE CORNEC N.D.R.

## INTERNATIONAL SEARCH REPORT

onal Application No  
PCT/US 01/27628

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	abstract	
A	<p>--- WO 97 33551 A (MILLENNIUM PHARMACEUTICALS) 18 September 1997 (1997-09-18) the whole document</p>	1-60
A	<p>--- WO 00 20869 A (RAULAI DANIEL ;SOUTTOU BOUSSAD (FR); VIGNY MARC (FR); UNIV GEORGET) 13 April 2000 (2000-04-13) the whole document</p>	1-60
A	<p>--- WO 99 67386 A (CHIRON CORP) 29 December 1999 (1999-12-29) the whole document</p>	1-60
L	<p>--- DATABASE EMBL [Online] 26 June 2001 (2001-06-26) T. OTA ET AL: "Human cDNA sequence seq ID no.14325" retrieved from EBI, HINXTON, UK Database accession no. AAH15833 XP002219862 L document cited to provide information on the relevant sequence ID no.14325 disclosed in EP1074617 the whole document</p>	1
P,X	<p>-&amp; EP 1 074 617 A (HELIX RESEARCH INSTITUTE) 7 February 2001 (2001-02-07)</p>	
A	<p>--- TOMMERUP N ET AL: "Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders" GENOMICS, ACADEMIC PRESS, SAN DIEGO, US, vol. 27, no. 2, 20 May 1995 (1995-05-20), pages 259-264, XP002117526 ISSN: 0888-7543 the whole document</p>	1-60
X,P	<p>--- DATABASE EMBL [Online] 29 September 2000 (2000-09-29) T. ISOGAI ET AL: "NEDO human DNA sequencing project. Homo sapiens cDNA FLJ12533 fis, clone NT2RM4000202, weakly similar to ZINC Finger PROTEIN MOK-2" retrieved from EBI, HINXTON, UK Database accession no. AK022595 XP002219863 abstract &amp; UNPUBLISHED,</p>	1
	<p>--- -/--</p>	



In International Application No  
PCT/US 01/27628

PCT/US 01/27628

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
L  P,X	DATABASE EMBL [Online] 26 June 2001 (2001-06-26) T. OTA ET AL: "Human protein sequence ID no.14326" retrieved from EBI, HINXTON, UK Database accession no. AAB94103 XP002219864 L document cited to provide information on the relevant sequence ID no.14326 disclosed in EP1074617 the whole document -& EP 1 074 617 A (HELIX RESEARCH INSTITUTE) 7 February 2001 (2001-02-07) -----	1

## INTERNATIONAL SEARCH REPORT

national application No.  
PCT/US 01/27628**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
  
Although claims 56 and 59 are directed to a method of treatment of the human/animal body (rule 39.1 IV PCT), the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 55, 56, 58, 59 all partially  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
  
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
  
1-60 (all partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 55, 56, 58, 59 all partially

Claims 55-56 refer respectively to a composition comprising an agonist compound of the MDDT polypeptide and to a method for treating a disease using said composition, claims 58-59 refer respectively to a composition comprising an antagonist compound of the MDDT polypeptide and to a method for treating a disease comprising administering said composition without giving a true technical characterization. Moreover, no such compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported.

No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved. But a partial search has been carried out as far as the agonist/antagonist compound is an antibody against the MDDT polypeptide.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

**FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210**

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

**Invention 1: claims 1-60 all partially**

Polynucleotide represented by the polynucleotide sequence ID no. 1 and protein encoded thereby represented by sequence ID no.253. Antibodies against said protein, nucleic acid probes derived from said nucleic acid. Method of detection, treatment, diagnosis, screening and therapeutic compositions using said protein, antibodies and/or nucleic acids.

**Invention 2: claims 1-60 all partially**

Polynucleotide represented by the polynucleotide sequence ID no. 2 and protein encoded thereby represented by sequence ID no.254. Antibodies against said protein, nucleic acid probes derived from said nucleic acid. Method of detection, treatment, diagnosis, screening and therapeutic compositions using said protein, antibodies and/or nucleic acids.

**Inventions 3-252: claims 1-60 all partially**

Polynucleotide represented by the polynucleotide sequence ID no.3 to no.252 and protein encoded thereby represented by its corresponding sequence ID. Antibodies against said protein, nucleic acid probes derived from said nucleic acid. Method of detection, treatment, diagnosis, screening and therapeutic compositions using said protein, antibodies and/or nucleic acids.

## INTERNATIONAL SEARCH REPORT

Information on patent family members

In International Application No  
PCT/US 01/27628

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9733551	A	18-09-1997	AU 2582897 A	01-10-1997
			US 6268470 B1	31-07-2001
			US 6100031 A	08-08-2000
			WO 9733551 A2	18-09-1997
			US 6458939 B1	01-10-2002
-----				
WO 0020869	A	13-04-2000	CA 2346406 A1	13-04-2000
			EP 1121598 A1	08-08-2001
			JP 2002526777 T	20-08-2002
			WO 0020869 A1	13-04-2000
-----				
WO 9967386	A	29-12-1999	US 6262249 B1	17-07-2001
			AU 4705199 A	10-01-2000
			EP 1090117 A2	11-04-2001
			WO 9967386 A2	29-12-1999
			US 2001016651 A1	23-08-2001
-----				
EP 1074617	A	07-02-2001	AU 6180800 A	19-02-2001
			AU 6180900 A	19-02-2001
			AU 6181000 A	19-02-2001
			AU 6181100 A	19-02-2001
			AU 6181200 A	19-02-2001
			AU 6181300 A	19-02-2001
			AU 6181400 A	19-02-2001
			AU 6181500 A	19-02-2001
			AU 6181600 A	19-02-2001
			AU 6315800 A	19-02-2001
			EP 1074617 A2	07-02-2001
			EP 1205549 A1	15-05-2002
			WO 0109315 A1	08-02-2001
			WO 0109345 A1	08-02-2001
			WO 0109316 A1	08-02-2001
			WO 0109349 A1	08-02-2001
			WO 0109317 A1	08-02-2001
			WO 0109318 A1	08-02-2001
			WO 0109319 A1	08-02-2001
			WO 0109346 A1	08-02-2001
			WO 0109320 A1	08-02-2001
			WO 0109321 A1	08-02-2001
			WO 0109322 A1	08-02-2001
			WO 0109323 A1	08-02-2001
			JP 2002171977 A	18-06-2002
			JP 2002191363 A	09-07-2002
			US 2003017480 A1	23-01-2003
-----				